An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action (2025)

DETAILS

206 pages | 6 x 9 | PAPERBACK

ISBN 978-0-309-09895-3 | DOI 10.17226/29087

CONTRIBUTORS

Laura Adams, Elaine Fontaine, Michael Matheny, Sunita Krishnan, Editors; The Learning Health System Series; National Academy of Medicine

SUGGESTED CITATION

National Academy of Medicine. 2025. *An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action*. Washington, DC: The National Academies Press. https://doi.org/10.17226/29087.

THE LEARNING HEALTH SYSTEM SERIES

An Artificial Intelligence Code of Conduct for Health and Medicine

Essential Guidance for Aligned Action

Laura Adams, Elaine Fontaine,
Michael Matheny, Sunita Krishnan, ***Editors***

NATIONAL ACADEMY of MEDICINE

WASHINGTON, DC
NAM.EDU

NATIONAL ACADEMIES PRESS
Washington, DC

NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

This publication has undergone peer review according to procedures established by the National Academy of Medicine (NAM). Publication by the NAM signifies that it is the product of a carefully considered process and is a contribution worthy of public attention but does not constitute an endorsement of conclusions and recommendations by the NAM. The views presented in this publication are those of individual contributors and do not represent formal consensus positions of the authors' organizations; the NAM; or the National Academies of Sciences, Engineering, and Medicine.

This publication was completed with support from the Gordon and Betty Moore Foundation, the California Health Care Foundation, the Patrick J. McGovern Foundation, and the National Institutes of Health (Contract Number HHSN263201800029I). Any opinions, findings, or conclusions expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for this publication.

International Standard Book Number-13: 978-0-309-XXXXX-X
Digital Object Identifier: https://doi.org/10.17226/29087
Library of Congress Catalog Number: XXXXXXXXX

Printed in the United States of America.

Suggested citation: National Academy of Medicine. 2025. *An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action.* L. Adams, E. Fontaine, M. Matheny, and S. Krishnan, editors. NAM Special Publication. Washington, DC: National Academies Press. https://doi.org/10.17226/29087.

"Knowing is not enough; we must apply.
Willing is not enough; we must do."
—GOETHE

LEADERSHIP
INNOVATION
IMPACT

for a healthier future

NATIONAL ACADEMY *of* MEDICINE

ABOUT THE NATIONAL ACADEMY OF MEDICINE

The **National Academy of Medicine** is one of three Academies constituting the National Academies of Sciences, Engineering, and Medicine (the National Academies). The National Academies provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on issues of health, health care, and biomedical science and technology. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

Learn more about the National Academy of Medicine at NAM.edu.

AUTHOR GROUP

ANDREW BINDMAN, Kaiser Permanente
GRACE CORDOVANO, Enlightening Results
JODI DANIEL, Crowell Health Solutions
WYATT DECKER, UnitedHealth Group
PETER J. EMBÍ, Vanderbilt University Medical Center
GIANRICO FARRUGIA, Mayo Clinic (*Author Group Co-Lead*)
KADIJA FERRYMAN, Johns Hopkins University
SANJAY GUPTA, Emory University
ERIC HORVITZ, Microsoft
ROY JAKOBS, Royal Philips (*Author Group Co-Lead*)
KEVIN B. JOHNSON, University of Pennsylvania
PETER LEE, Microsoft
KENNETH MANDL, Harvard University
KEDAR MATE, Institute for Healthcare Improvement
DEVEN McGRAW, Citizen Health
BAKUL PATEL, Google (*Author Group Co-Lead*)
PHILIP PAYNE, Washington University
VARDIT RAVITSKY, The Hastings Center
SUCHI SARIA, Bayesian Health and Johns Hopkins University
ERIC TOPOL, Scripps Research
SELWYN VICKERS, Memorial Sloan Kettering Cancer Center

Development of this publication was facilitated by the contributions of the following people:

Editors:

LAURA ADAMS, National Academy of Medicine
ELAINE FONTAINE, National Academy of Medicine
SUNITA KRISHNAN, National Academy of Medicine
MICHAEL MATHENY, Vanderbilt University Medical Center

Contributors:

DAVID DORR, Oregon Health & Science University
ANDREA DOWNING, Light Collective
TYLER LOFTUS, University of Florida
SHAUNA OVERGAARD, Mayo Clinic
RAVI B. PARIKH, Emory University

Staff

Development of this publication was facilitated by contributions of the following NAM staff, under the guidance of J. Michael McGinnis, Leonard D. Schaeffer Executive Officer and Executive Director of the NAM Leadership Consortium:

LAURA ADAMS, Senior Advisor
AUDREY ELLIOTT, Associate Program Officer
SUNITA KRISHNAN, Senior Program Officer
ANNIE MURFF, Senior Program Assistant

DISCLOSURE OF CONFLICTS OF INTEREST

Authors, editors, and contributors to this publication have filed disclosure statements on perceived or actual interests related to the development and application of artificial intelligence in health, health care, and biomedical science. These statements cover the following circumstances:

1. ORGANIZATIONAL AFFILIATIONS. Business relationships (as an employee, owner, officer, director, consultant, etc.) and remunerated or volunteer non-business relationships (e.g., professional organizations, trade associations, public interest or civic groups) that might be relevant to this National Academy of Medicine (NAM) activity.
2. FINANCIAL INTERESTS. Significant financial interests (research support, investments, intellectual property, commercial holdings) that might be perceived as relevant to this NAM activity.
3. ASSOCIATES. Significant interest known to you as part of close associates or friends whose welfare might be viewed as relevant to this NAM activity.
4. FAMILY MEMBERS. Significant interests of family members whose welfare might be viewed as relevant to this NAM activity.

A summary of relevant items is included with the author information starting on p. 171.

See Author Information on p. 171.

See Editor Information on p. 182.

See Contributor Information on p. 184.

ACKNOWLEDGMENTS

The following individuals participated in expert working groups to consider and characterize real-world activities and stakeholder perspectives, responsibilities, and needs in the context of the Artificial Intelligence Code of Conduct.

PAT BAIRD, Philips
SURESH BALU, Duke University
LISA BARI, Civitas Networks for Health
VALERIE BONHAM, Kennedy Krieger Institute
ANASTASIA CHRISTIANSON, Pfizer
THERESA CULLEN, Pima County Health Department
MOHAMMAD HOSSEINI, Northwestern University
TIMOTHY HSU, Association for the Advancement of Medical Instrumentation
ABEL KHO, Northwestern Medicine
DAVID MARC, Community Clinical Services
GENEVIEVE MELTON-MEAUX, University of Minnesota
STEVE MIFF, Parkland Center for Clinical Innovation
RENE QUASHIE, CTA
MARK SENDAK, Duke University
KAVEH SHOJANIA, University of Toronto
WILL SHRANK, Andreessen Horowitz
LAUREN SILVIS, Tempus
KARANDEEP SINGH, University of California, San Diego
MARINA SIROTA, University of California, San Francisco
BOB WACHTER, University of California, San Francisco
NICOLE WEISKOPF, Oregon Health & Science University
SHANNON WEST, Datavant

REVIEWERS

This Special Publication was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academy of Medicine in making its published Special Publication as sound as possible and to ensure that the Special Publication meets the institutional standards for quality, objectivity, and evidence. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We wish to thank the following individuals for their contributions:

DAVID BLUMENTHAL, Harvard University
VINCENT LIU, Kaiser Permanente
RAJ RATWANI, MedStar Health Research Institute
EDWARD SHORTLIFFE, Columbia University
WILLIAM STEAD, Vanderbilt University

The reviewers listed above provided many constructive comments and suggestions, but they were not asked to endorse the content of the Special Publication and did not see the final draft before it was published.

Review of this publication was overseen by **PAUL C. TANG,** Stanford University School of Medicine. Responsibility for the final content rests entirely with the author group and the National Academy of Medicine.

CONTENTS

BOXES, FIGURES, AND TABLES

BOXES

FIGURES

TABLES

PREFACE

Recent advancements in artificial intelligence (AI) technologies have unlocked unprecedented opportunities in health, health care, and biomedical science, and these breakthroughs hold the potential to fundamentally transform approaches to medical and health research, health promotion, disease prevention, diagnosis, treatment, and health system management. By enabling more effective, efficient, and personalized care, AI stands poised to address some of the most persistent challenges in the health sector, provided it is properly governed and effectively stewarded. As the health sector grapples with converging challenges, AI has the potential to be a much-needed transformative force, capable of helping solve some of the most intractable issues in health care today. These issues—including inequities in access, rising costs, clinician burnout, and the growing burden of chronic disease—demand bold, new approaches. The promise of AI extends beyond its technological capabilities to encompass a more profound reconsideration of care delivery, aiming to improve outcomes for everyone, particularly our most vulnerable.

However, the development and deployment of AI introduces critical ethical, accountability, and safety considerations. As AI technologies diffuse into health care, it is vital to establish robust guidance for their integration, ensuring alignment with the foundational commitment to improve health and well-being for all.

This NAM Special Publication, *An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action*, addresses this imperative. By harmonizing existing AI principles, identifying gaps, and aligning them with the core commitments of the NAM's Learning Health System (LHS), it provides a comprehensive, adaptable set of guidelines for health care organizations and stakeholders. These principles are intended as guideposts for the development, implementation, and continuous improvement of AI systems, ensuring they uphold the highest standards of integrity, safety, and effectiveness.

As AI becomes more embedded in health care, it is essential to ensure that its use fosters trust and collaboration across the health care ecosystem. In the spirit of the NAM's Shared Commitments that set out expectations for all participants and stakeholders in health and health care, the Code Commitments presented in this publication offer a touchstone for organizations as they develop their AI strategies and approaches. By fostering collaboration and shared understanding, the Code of Conduct aims to minimize the risks of fragmented efforts and maximize the collective impact of AI in improving health care.

At the heart of this work is the belief that AI has the potential to play a pivotal role in creating a more effective, accessible, and sustainable health care system. This publication not only explores the role of this technology in advancing human health but also underscores the importance of ensuring that its benefits reach all individuals, creating a future where excellence in health care is a reality for everyone.

J. Michael McGinnis, MD, MPP
Leonard D. Schaeffer Executive Officer
National Academy of Medicine

Victor Dzau, MD
President
National Academy of Medicine

ACRONYMS AND ABBREVIATIONS

ADRD	automated diabetic retinopathy detection
AHRQ	Agency for Healthcare Research and Quality
AI	artificial intelligence
AICC	AI Code of Conduct
AMIA	American Medical Informatics Association
ASA	Assistant Secretary for Administration
ASTP ONC	Assistant Secretary for Technology Policy, Office of the National Coordinator for Health Information Technology
CDC	Centers for Disease Control and Prevention
CMS	Centers for Medicare & Medicaid Services
DNN	deep neural network
DR	diabetic retinopathy
EHI	electronic health information
EHR	electronic health record
EU	European Union
FDA	U.S. Food and Drug Administration
FM/GAI	foundational models/generative AI
FOA	funding opportunity announcement
HHS	Department of Health and Human Services
HITECH	Health Information Technology for Economic and Clinical Health
HRSA	Health Resources and Services Administration

IOM	Institute of Medicine
ISO	International Organization for Standardization
IT	information technology
KRB	knowledge and rule-based
LHS	Learning Health System
LLM	Large Language Model
ML	machine learning
NAHQ	National Association for Healthcare Quality
NAM	National Academy of Medicine
NAS	National Academy of Sciences
NEHRS	National Electronic Health Records Survey
NSF	National Science Foundation
OCR	Office for Civil Rights
OECD	Organisation for Economic Co-operation and Development
ONC	Office of the National Coordinator for Health Information Technology
REC	regional extension center
TEFCA	Trusted Exchange Framework and Common Agreement
TPB	Theory of Planned Behavior
UK	United Kingdom
UN	United Nations
WHO	World Health Organization

EXECUTIVE SUMMARY

In the past decade, significant advances in artificial intelligence (AI) technologies have created transformational opportunities for health, health care, and biomedical science, offering new tools to improve effectiveness and efficiency in a myriad of applications in the health field. Advancing AI capabilities are arriving contemporaneously to new and exacerbated challenges in health care, including staff burnout and shortages, an aging population with growing disease burden, increasing costs of care, and persistent issues with equity in access and outcomes. Despite sustained attention, conventional strategies have not yielded the desired improvements in population health, cost containment, patients' experience, clinician well-being, or equitable outcomes. The need for new approaches to address these long-standing challenges is evident; AI offers both new hope and new concerns.

AI is being used and championed to drive transformative progress in diagnostics, population health, care quality, patient safety, clinician experience, and clinical and administrative efficiency. AI risk prediction tools, such as cardiovascular risk calculators, have been in use for decades. AI use in radiology is widespread (Yordanova, 2024); AI tools are showing promise in improving speed and accuracy of pathology results (Greeley et al., 2024; McGenity et al., 2024); and AI tools are being used to automate clinical documentation and other routine tasks (Abdelhady and Davis, 2023). AI is increasingly being used and evaluated in clinical decision support applications to support more tailored and targeted recommendations. And generative AI interactive chatbots are being employed in a variety of health and health care settings (Kurniawan et al., 2024).

However, certain concerns about AI have also been raised. Much attention has been given to risks of exacerbation of existing system biases as well as introduction of new forms of bias (Obermeyer et al., 2019; Topol, 2019). An important issue related to equity is scope of access to the benefits of AI and the need to ensure availability to transformative tools in to both high- and low-resource health care settings. Additional concerns include data security and individual privacy; accuracy and reliability of results; explainability, transparency, and accountability; and influences on humans, whether through impacts of automation on the workforce or through influences of AI anthropomorphism on individuals and patients. Other challenges are related to user preferences and workflow integration, which are present in any health technology, but have distinct manifestations in health AI.

The integration of AI into health, health care, and biomedical science applications requires realization of AI's transformative potential. The ethical deployment of AI necessitates proactive harm-reduction strategies to ensure that its positive effects are maximized while minimizing negative externalities. To this end, many entities, from the organizational to the global level, have developed AI governance frameworks to address these concerns.

The objective of the AI Code of Conduct (AICC) project is to harmonize the existing principles, address identified gaps, and map these principles to the National Academy of Medicine's Learning Health System (LHS) Shared Commitments (McGinnis et al., 2024). From this set of aligned principles, a small number of simple rules were developed based on complex adaptive systems theory which posits that individuals adhering to a concise set of agreed-upon simple rules can create change at the system level (Adams et al., 2024). The AICC framework is deliberately designed as a touchstone for organizations and groups developing their own approaches and considerations to consider for inclusion and alignment when assessing internal guidance for completeness in their specific context, thereby advancing trust and minimizing the likelihood of actors across the field working at cross-purposes. This set of simple rules is presented as Code Commitments. Described in detail in Table ES-1, they are intended to serve as guideposts for all stakeholders as they develop and use AI.

Commitment 1: Advance Humanity

- Development of standards and other governance structures to assess alignment by developers and users of health AI with societal and cultural goals for health AI
- Incentives and structures for independent evaluation, certification to the AI Code Commitments, and public and transparent reporting on certification status

Commitment 2: Ensure Equity

- Standardized metrics to assess and report bias in data, AI output, and AI use, in the interest of equitable distribution of benefit and risk
- Incentives and supports to low-resourced organizations and communities to ensure equitable access to the benefits of AI

Commitment 3: Engage Impacted Individuals

- Participation by all key stakeholders across the health AI lifecycle
- Local governance bodies that include all stakeholders in the AI lifecycle cross-purposes
- Common understanding and education of all affected parties

Commitment 4: Improve Workforce Well-Being

- Positive work and learning environments and culture (NAM, 2022c)
- Measurement, assessment, strategies, and research (NAM, 2022c)
- Reskilling and training programs for workforce AI competency
- Disruptive technologies with change management strategies that promote worker well-being

Commitment 5: Monitor Performance

- Standardized quality and safety metrics to be used to assess the impact of the use of health AI on health outcomes
- Aligned frameworks for safety, equity, and quality in AI performance

Commitment 6: Innovate and Learn

- A well-supported national health AI research agenda
- Participation in shared learning across all stakeholders
- Innovation as a core investment

Key stakeholders—including developers, researchers, health systems and payors, patients, ethics and equity experts, quality experts, and former leaders of federal agencies—considered the role of their respective group in applying the Code Commitments. Section V of this publication details their work efforts. Within the main report, a table is provided that describes commonalities in stakeholder actions, and Table ES-1 describes elements in which one stakeholder group has distinguishing or leading roles and responsibilities. This includes important considerations such as ensuring broad stakeholder participation throughout the AI lifecycle, assessing AI tools initially and continuously for technical and health outcomes, promoting transparency, promoting continuous learning, identifying and addressing conflicts of interest and objectives, considering and addressing bias, implementing incentives to encourage desired behavior, promoting user-centered design, promoting local governance, and creating and supporting a culture of safety.

To achieve trustworthy health AI at scale, the development, implementation, and use and monitoring of and ongoing learning about health AI will require intentional and sustained collaboration between and among all impacted

TABLE ES-1 | Distinct Contributions of Various Stakeholders to the Application of the AI Code of Conduct Framework

Stakeholder	Distinct Contribution
Developers	Developers have vast experience in methods and practices, and their active participation in developing standards for the industry will be foundational.
Researchers	Researchers are positioned to provide scientifically sound and independent assessment of both methods and outcomes associated with health AI, including issues of data de-identification, and sharing and the associated implications of societal benefits and burdens, as well as the best practices and standards in workflow integration.
Health Systems and Payors	Local adaptation that facilitates human agency and promotes patient-centered care is the purview of health systems and payors, as is the training and support of the health care workforce in the use of AI in the local health care delivery context. They have an opportunity to create financial incentives that support equitable and effective health AI, using both increases and decreases in reimbursement to support desired best practices around AI use.
Patients and Advocates	Patients, as the recipients of health AI, are uniquely positioned to describe in detail their experience about the impacts of health AI on their lives. Only they can articulate their preferences about critical issues such as access controls over their data or explanations about when and how AI is used in their care. Only patients can share their own personal experiences, both positive and negative, of engagement with developers and the health care system as the use of AI for diagnosis, treatment, and payment advances. And patients are by definition the only source of patient-reported outcomes.
Federal Agencies	The funding and regulatory authority held by the federal agencies has the power to shape the future of health AI. Some examples of how these tools could take form include through support for studies to measure how AI can influence patient health, human agency, goals of care, and human-human interactions in the presence of AI interventions; through recognition of standards for collection and exchange of relevant data and encouraging use of the Trusted Exchange Framework and Common Agreement for making data available for training algorithms, and prioritized research projects; or through the expansion of requirements in AI product labeling based on real-world performance.
Health Care Workforce	As end-users of some types of health AI, the health care workforce is situated to identify workflow needs and priorities, and as purchasers or influencers of purchasing decisions, clinicians in particular may have contracting opportunities to require disclosure of AI models' alignment with the Code principles and commitments and to address liability concerns should model outputs cause harm.
Quality and Patient Safety Experts	Quality and patient safety experts and accrediting organizations play the role of an independent auditor, ensuring that processes are designed and implemented and that metrics are developed and routinely assessed to ensure the quality of outputs and reduce the risk of harm from health AI tools.
Ethicists and Equity Experts	Ethics and equity experts are uniquely qualified to consider and weigh the novel tensions that health AI presents across various stakeholder priorities, always holding the greatest good for the health of the individual and the community as the north star. They are positioned to serve as guides on the path to implementing trustworthy AI.

stakeholders. Herein, the Tight-Loose-Tight leadership model (see Figure ES-1) is applied as a construct to scale change in complex systems; it is designed to balance innovation and control through an iterative, dynamic approach that encourages collaboration and builds trust. During the first Tight phase, the goal is to align stakeholders on vision and expectations; an example is alignment on governance frameworks. During the Loose phase, local implementation leads to learning and innovation. Finally, the second Tight phase promotes change at scale through evaluation metrics, and standards, among others. Key components to advance health AI through the Tight-Loose-Tight framework are presented in Table ES-2.

Finally, a small number of strategic priorities were identified by the authors as most likely to support the achievement of each Code Commitment. These key priorities with associated actions and responsible actors are described in Table ES-2 by Code Commitment:

With intentional, sustained effort and ongoing communication, feedback, and collaboration by all stakeholders, safe, effective, and efficient advancement of responsible health AI is possible. Realizing the benefits and mitigating the

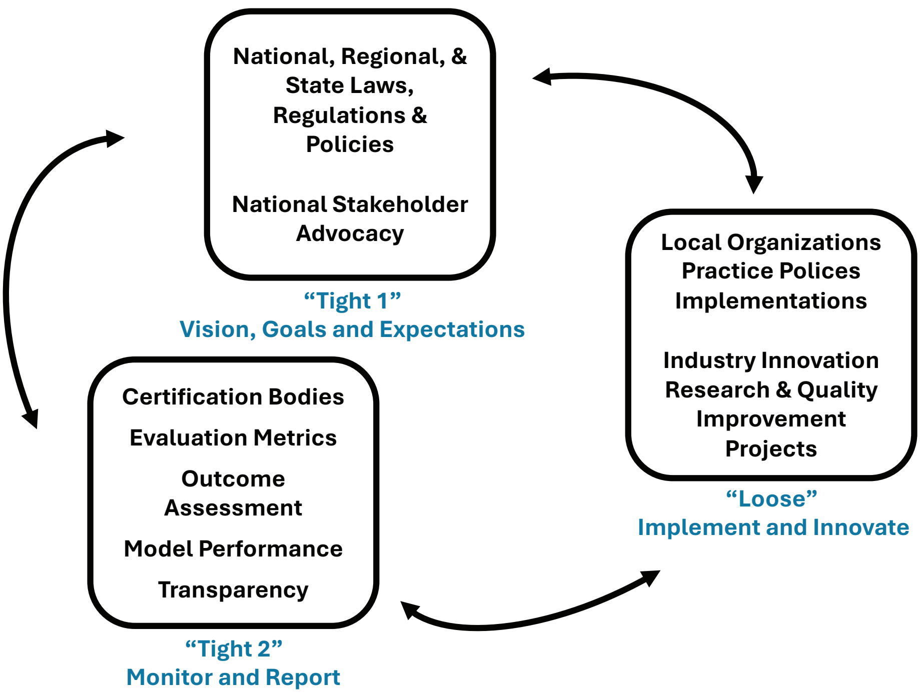

FIGURE ES-1 | Representation of the Tight-Loose-Tight model of leading change at scale adapted for the AI in health and health care context.
SOURCE: Conceptually adapted from Compton-Phillips, 2019.

TABLE ES-2 | Summary Priority Actions to Operationalize the AICC Code Commitments

Commitment	Action	Involved Parties
Advance Humanity	• Development of governance standards for AI alignment with societal goals. • Incentives and structures for independent evaluation, certification to the AI Code Commitments, and public reporting.	• Developers; health systems and payors; researchers; ethicists; professional associations, state, federal, and international governments; patients, families, and communities • Federal agencies including ASTP ONC, FDA, NIH
Ensure Equity	• Standardized metrics to assess and report bias in data, AI output, and AI use, in the interest of equitable distribution of benefit and risk. • Incentives and supports to low-resourced organizations and communities to ensure equitable access to the benefits of AI.	• Researchers and federal agencies • Federal agencies including ASTP ONC, FDA, NIH, HRSA
Engage Impacted Individuals	• Participation by all key stakeholders across the health AI lifecycle. • Local governance bodies that include all stakeholders in the AI lifecycle. • Common understanding and education of all impacted parties.	• Federal agencies including ASTP ONC, FDA, NIH, HRSA • Developers; health systems and payors; researchers; ethicists; professional associations, state, federal, and international governments; patients, families, and communities • Developers; health systems and payors; researchers; ethicists; professional associations, state, federal, and international governments; patients, families, and communities
Improve Workforce Well-Being	• Positive work and learning environments and culture (NAM, 2022c). • Measurement, assessment, strategies, and research (NAM, 2022c). • Reskilling and training programs for workforce AI competency. • Disruptive technologies with change management strategies that promote worker well-being.	• Developers, health systems and payors, researchers • Developers, health systems and payors, researchers, and federal agencies (e.g., NIH) • Researchers, educational institutions, federal agencies (e.g., Department of Education), professional societies • Health systems and payors

TABLE ES-2 | Continued

Commitment	Action	Involved Parties
Monitor Performance	• Standardized quality and safety metrics to assess the impact of the use of health AI on health outcomes. • Aligned frameworks for safety, equity, and quality in AI performance.	• Federal agencies, researchers, accreditation bodies, patient safety organizations • Federal agencies, researchers, accreditation bodies, patient safety organizations
Innovate and Learn	• A well-supported national health AI research agenda. • Participation in shared learning across all stakeholders. • Innovation as a core investment.	• Federal agencies (e.g., NIH) and researchers • Federal agencies (e.g., ASTP ONC)

NOTE: ASTP ONC = Assistant Secretary for Technology Policy, Office of the National Coordinator for Health Information Technology; FDA = U.S. Food and Drug Administration; HRSA = Health Resources and Services Administration; NIH = National Institutes of Health.

risks will require significant engagement, which will be more likely to come to fruition if it is easy and rewarding to abide by the shared vision, values, goals, and expectations described in the nationally aligned AI Code Principles and Code Commitments.

1
INTRODUCTION

As artificial intelligence (AI) methods and capabilities have advanced over the past decade, interest in the application of AI to make meaningful contributions to disease prevention, management, and cure has surged. These advances in AI come at an opportune moment for health care as the field struggles with significant challenges, both long-standing and exacerbated by the COVID-19 pandemic. Concerns include but are not confined to rising costs, strained staff and staffing, limited and inequitable access, disparities in outcomes, growing disease burden, and persistent patient safety challenges. According to the most recent data available from the Centers for Medicare & Medicaid Services, national spending on health care reached 17.3% of gross domestic product in 2022 (CMS, 2024). A recent literature review reflected that 35% of physicians are experiencing burnout (Hoff et al., 2023). Nearly half of adults surveyed by the Kaiser Family Foundation through ongoing polling, updated in March 2024, reported difficulty in affording health care (Lopes et al., 2024). Disparities in care and outcomes persist as documented by a recent analysis of maternal outcomes that demonstrated that Black mothers in the top 20% of income distribution experienced a mortality rate of 4.3 deaths per 10,000, while White women in the lowest income quintile had a maternal mortality rate of 2.7 deaths per 10,000, a difference of nearly 60% that cannot be explained by access and resources (Kennedy-Moulton et al., 2023). Estimates from the Centers for Disease Control and Prevention indicate that the prevalence of diabetes in the U.S. population continues its upward trend, growing from 10.3% in the period between 2001 and 2004 to 13.2% in the period between 2017 and 2020 (CDC, 2024). And, since the publication of the *To Err Is Human* (IOM, 2000) describing the staggering number of deaths each year due to medical errors, progress in improving patient safety has been inadequate (Bates and Singh, 2018). Clearly opportunities for improvement abound.

In the last several years, AI tools have shown the potential to ameliorate many of these systemic challenges (e.g., improved accuracy and efficiency, cost reduction, and staff augmentation) and may also play a role in truly solving them (Cutler, 2023). While the health care field is still at the beginning of integrating AI capabilities into clinical practice and administration, early tools and experimentation have yielded promising, and in some cases outstanding, results. The most mature efforts in integrating AI in clinical care may be in radiology where it has been employed across almost all subspecialties and modalities over the past decade (Yordanova, 2024). There is intense interest and investment in AI-powered products using ambient listening for clinical notetaking, which show promise but are not yet established (Tierney et al., 2024). Meanwhile, AI-powered tools have proven effective in performing many administrative tasks by automating documentation and other routine tasks (Abdelhady and Davis, 2023). AI is also streamlining complex health care supply chains by getting the right resources to the right people faster. For example, in Ghana and Rwanda, AI-enabled drones are delivering life-saving vaccines, medications, and blood products to remote and underserved regions (Krittanawong and Kaplin, 2021). By leveraging AI tools trained across interconnected, highly secure, standardized and de-identified datasets, researchers, clinicians, and innovators are exploring how AI can both improve the current care delivery system and create opportunities for improved patient outcomes (Halamka and Cerrato, 2021). Indeed, as AI continues to improve and mature, it holds promise to impact many long-standing challenges and to create new opportunities to markedly improve health care delivery and health outcomes.

Administration, logistics, patient navigation, preventive care, clinical care, and virtual care modalities are all poised for marked improvements due to cutting-edge AI applications. For example, AI-powered tools show great promise to increase the speed and accuracy of pathology results interpretation (Greely et al., 2024; McGenity et al., 2024). In direct clinical care, AI-assisted procedures—such as AI-assisted colonoscopies (Wallace et al., 2022) that promote fuller examination of the colon—have significant potential to support higher quality of care delivery and improvement in health outcomes for patients across the world, resulting in less disease and fewer complications for more people. AI models are being used to predict risk of clinical deterioration in patients in intensive care units, aiding in more rapid interventions and resulting in decreased mortality (Escobar et al., 2020). AI-enabled devices and systems, running continuously as seamless end-to-end solutions, may allow both patients and their care teams to detect and prevent diseases earlier than conventional testing—such as through AI-enabled ECG analysis that when accompanied by leadership support, show promise to identify early heart failure before symptoms develop (Yao et al., 2021). AI has also shown the potential to transform personalized medicine, making care faster, more precise, and more efficacious, such as algorithms that analyze genomic data to precisely anticipate a patient's response to specific therapeutic interventions—allowing patients to receive more effective and potentially safer treatments (Tao et al., 2020). Moreover, generative AI shows early promise in simplifying pathology reports for patients navigating the health care system by explaining test results, outcomes, or billing (Steimetz et al., 2024) and acting as a kind of interactive health care "encyclopedia" (Reddy, 2024) and may be capable of providing medical language interpretation in the future.

The overwhelming majority of individual health is managed through self-care and caregiving in the community, outside of the traditional health care delivery system. Patients and caregivers are increasingly utilizing personal and health care certified AI tools, applications, and devices to support self-management. Correspondingly, health care research and operations personnel are developing new tools to collect and analyze these important health data. AI-enabled wearables and remote monitoring tools that assess sleep, breathing, cardiac rhythms, and ambient voice data (Bajwa et al., 2021) have the potential to support self-care and to connect patient experiences more closely than ever to medical research, helping to close the gaps that exist in delivering evidence-based medicine.

There is great interest in using available AI tools to improve business operations and clinical care; furthermore, AI experts anticipate that ongoing scientific advances will lead to AI-enabled capabilities that are currently not possible (Anderson and Rainie, 2023). AI-assisted health research is exploring new capabilities in the biomedical sciences that accelerate drug discovery and design, identify novel targets, and help assess toxicity (Mullowney et al., 2023). Clearly, the potential for AI to improve health and health care is substantial, as is investment—both public and private—in AI research to drive the development of new tools and techniques. In 2021, U.S. non-defense agencies allocated $1.5 billion to AI research, while the private sector spent more than $340 billion (Ahmed et al., 2023b). The drive to continue to innovate is creating competition across the industry and the world (O'Brien, 2024; YOLE Group, 2024), and in the current unregulated environment, the incentives for personal gain could stymie the public good.

Concomitant to the drive and investment to rapidly advance AI techniques, tools, and adoption, there are important challenges and limitations that must be addressed to fully realize the promise and potential of AI in the health sector. Some examples include risks to safety, security, equity, and accountability. Additional existential threats such as misinformation, job loss, and widespread surveillance may be introduced by AI systems (Anderson and Rainie, 2023). From the perspective of patient safety and effectiveness of care, failure of an AI tool to accurately predict target outcomes could result in care that is of lower quality than care delivered without AI support. For example, a model that inaccurately predicts a negative patient condition (e.g., infection) in an acute care setting could lead to delays in care. Such a situation was reported when a validation study of an inpatient sepsis prediction model embedded in a widely used electronic health record demonstrated that, in practice, the model resulted in poor discrimination, missing some cases and over-alerting in others (Wong et al., 2021). The extremely large datasets used by AI also present privacy and security risks to patients. In 2024, multiple extremely large health data breaches were reported to the Office for Civil Rights, exposing data on hundreds of millions of individuals to nefarious actors (OCR, n.d.).

Health disparities and equity concerns can be created or exacerbated by AI. For example, the choice of data to include in a model can inadvertently build bias into the model. In the review of a model used to guide provision of resources to patients in need of supplemental health supports, Black patients were identified for additional care at a rate half that of White patients who were not as sick; this was due to the inclusion of health care costs as a proxy for illness burden in the model (Obermeyer et al., 2019). Another way in which inequities may be introduced is in access to the benefits of AI. Highly resourced organizations, such as academic medical centers, may be more able to implement new AI tools (Wu et al., 2023) than smaller, more poorly resourced organizations, such as federally qualified health centers, furthering the digital divide.

With the rapidly expanding use of AI in the health sector, accountability for development, testing, maintenance, and use also presents challenges that require careful attention (Davis et al. 2024). Newer AI methods may make transparency and explainability more difficult (Saeed and Omlin, 2023), reducing trust in the systems. Yet there are currently no national standards for assessing health care AI models, their outputs, or their governance, from their development and training to the necessary vigilant monitoring post implementation, to the outcomes they produce.

To realize the benefits, the risks associated with health AI must be addressed to ensure both high-quality, evidence-based care and to promote ongoing trust in the health system. The National Academy of Medicine's (NAM's) AI Code of Conduct (AICC) framework, presented below, was developed with guidance from the AICC steering committee and is designed to align the health field and to catalyze collective action to ensure that the transformative potential of AI to advance health, health care, and biomedical science is realized, while upholding the "highest standards of ethics, equity, privacy, security, and accountability" (NAM, n.d.). The AICC steering committee developed for public comment a draft AICC framework (Adams et al., 2024), comprising a set of Code Principles, informed by a literature review, harmonized with national and international guidance, and mapped to the NAM's Learning Health System (LHS) Shared Commitments (McGinnis et al., 2024). In the spirit of the NAM's Shared Commitments, the Code Commitments represent a set of simple rules or expectations intended to aid in the development and application of AI in health and health care. This publication represents a continuation of that work and delves deeper into the opportunities and challenges presented by health AI, presents updated Code Principles and Code Commitments, considers the requisite new actions and collaborations for impacted parties, and poses a set of priority actions for consideration by the various actors in the health arena to realize the vision of an LHS enabled by AI and guided by the Code of Conduct framework.

2
BACKGROUND

DEFINING ARTIFICIAL INTELLIGENCE

The term "artificial intelligence" (AI) was first introduced by John McCarthy at the seminal 1956 Dartmouth Conference, where the vision of making "machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves" was articulated (McCarthy et al., 1955). Overall, AI refers to a subdiscipline of computer science and a constellation of technologies that perform tasks that have traditionally required human intelligence. Core AI research areas include methods for learning, reasoning, problem-solving, planning, language and speech understanding, and visual perception.

Howell and colleagues (2024) suggest three "epochs of AI." In the first epoch, early AI was focused on symbolic and probabilistic reasoning (e.g., expert systems applied to decision pathways, Bayesian models used for clinical decision support). In medicine in the late 1950s, AI efforts centered on providing support for diagnosis and decision making, based on a foundation of probability and utility (Ledley and Lusted, 1959). In the second epoch, endeavors explored the use of logic-based inference (Buchanan and Shortliffe, 1984). Research in the 1980s sparked a statistical revolution in AI founded on principles of probability and utility that provide a frame for today's efforts and methods (Horvitz et al., 1988); this included advances in AI-powered medical diagnosis with rich representations known as Bayesian networks and related representations of probability and utility (Heckerman et al., 1992; Pearl, 1988).

In the last 20 years, AI has transitioned from a niche scientific pursuit to a foundational set of technologies poised to have impact in multiple domains, including health, health care, and biomedical science (NAM, 2022a). In the early 2010s, deep learning methods emerged and allowed programs to dramatically improve data-driven classification and prediction tasks (e.g., identification of diabetic retinopathy in retinal images (Gulshan et al., 2016). In the third and most recent "epoch of AI," foundational models and generative AI (including Large Language Models [LLMs]) have further extended deep learning methods, changing the paradigm from task-specific tools to tools that "can do many different things without being retrained" (Howell et al., 2024) (e.g., chatbots that can interact with patients for varied purposes). Taken together, recent advances in data availability, computing power, and computational methods have resulted in rapidly accelerating innovation (Horvitz and Mitchell, 2024). As noted above, advances in AI in recent years in data-driven machine learning (ML) techniques were foundational for these advances. ML-powered advances have been particularly impactful in the areas of language and vision. The advances with ML were also deeply dependent on the voluminous increases in data made available as a result of the massive digitalization of processes and communications as well as the pooling of such data made possible with broad interconnectivity, along with the dramatic decrease in storage costs. Additionally, ongoing specialization in computer chips (e.g., from central processing units, designed for general-purpose, sequential processing, to graphic processing units capable of high-speed parallel processing) have enabled the more intense computational demands of advanced ML procedures (Dean, 2022).

Several forms of data-driven ML have been employed in medicine. These include supervised, unsupervised, reinforcement learning, self-supervised learning, and newer techniques using deep neural networks. Figure 2-1 provides an overview of the categories and basic characteristics of both knowledge- and rule-based systems (noted above) and data-driven ML.

Supervised learning relies on labeled datasets, where the input data (e.g., findings and symptoms presented by a patient) are paired with the desired output (e.g., diagnoses). This approach has been used to train ML models to predict medical outcomes based on evidence presented to the AI systems. Supervised ML has been the basis for advances in performing medical diagnosis and predicting outcomes (Bayati et al., 2014; Wiens and Shenoy, 2018) and image analysis. Supervised ML has also been used to perform diagnoses from radiological and photographic imagery (e.g., skin, pathological sections, blood smears).

Unsupervised learning involves analyzing data without pre-defined labels. The goal is to find hidden patterns or intrinsic structures within the data. A common technique would be clustering. This type of learning is particularly useful for exploratory data analysis, such as grouping similar patients together based on presentation patterns (Alizadeh et al., 2000).

Self-supervised learning is a subset of unsupervised learning that generates its own labels from the data. It typically involves tasks such as predicting missing parts of

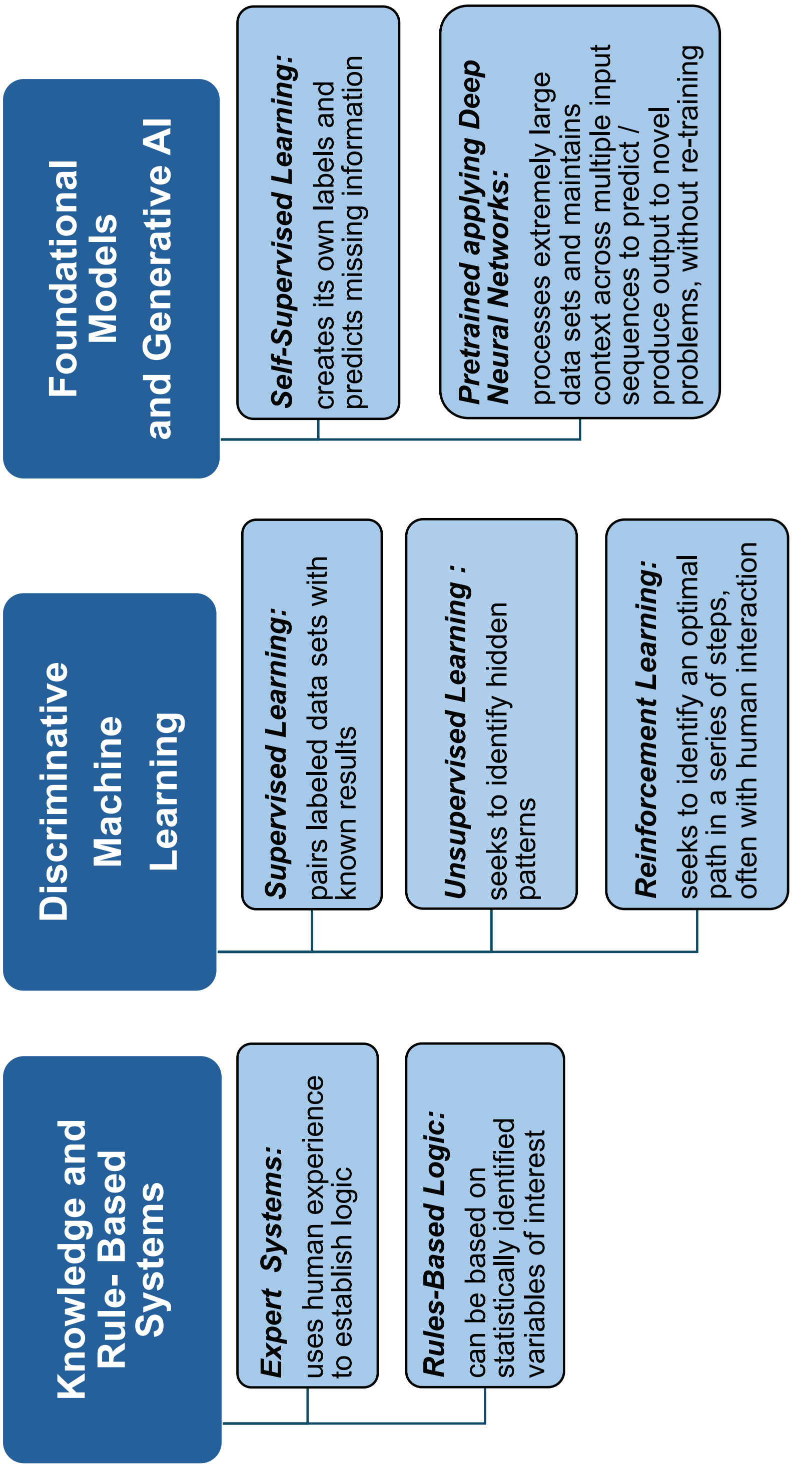

FIGURE 2-1 | High-level categories of AI.

the data. For example, a model might learn to predict the next word in a sentence or the next frame in a video sequence based on the context provided by previous words or frames. This method is a foundational training mechanism behind recent advances in natural language analysis and vision-based AI systems as it enables the training of models using vast amounts of unlabeled data.

Reinforcement learning is a method in which the purpose is for an agent to learn the optimal action path toward a stated goal or "reward" using experience in an environment rather than labeled datasets. Generally, these algorithms optimize for a long-term reward maximization based on the Markov decision process that recommends actions at discrete time steps based on this framework. This type of learning is particularly useful for recommendations based on user interactions and optimization challenges (Komorowski et al., 2018). An example of this would be to help decide on optimal treatment pathway recommendations in complex cancer treatments and surgical approaches, to maximize the probability of an outcome that should be prioritized, such as quality of life or lifespan (Khezeli et al., 2023).

On another dimension, ML can be defined by target uses and goals of the system. The uses can be broadly divided into *discriminative* and *generative* models, each with distinct objectives and application categories. *Discriminative models* classify items based on input features, such as medical diagnoses, laboratory tests, and wearables' data, mapping inputs to outcomes of interest. *Generative models*, on the other hand, learn from the patterns and relationships in training datasets, enabling the creation of new data instances that resemble training data. Generative models power AI applications in language generation, image creation, and scientific simulations.

Discriminative ML is highly promising for high-stakes medical decision support, as the methods are founded in well-characterized approaches to learning and reasoning under uncertainty. A few examples of this type of AI have been used in clinical practice widely as early examples were adapted to be calculated and used manually by clinicians (Bayati et al., 2014; Gomes et al., 2022; Mumtaz et al., 2023; Rajaguru et al., 2022; Xing et al., 2018).

Deep neural network (DNN) models—the development for which the 2024 Nobel Prize in Physics was awarded "for foundational discoveries and inventions that enable ML with artificial neural networks" (Nobel Prize Outreach AB, 2024)—are based on the automated construction of multilayered computational neural networks (often compared to the structures in the brain in their layered connectivity) that have propelled AI capabilities forward. This approach used in models, such as DeepFold (Pearce et al., 2022) and AlphaGo (Kemmerling et al., 2024), may hold promise if applied to various challenges in health research and care delivery. AI is currently experiencing an inflection point, driven by DNNs over the last decade. Great amounts of data and advances in computer power coupled with core algorithmic innovations (e.g., back-propagation and convolution), and developments with larger computing architectures (e.g., Transformers) have contributed to AI success in tasks such as speech recognition and image analysis. Significant progress has been made possible by DNNs, including reducing word error rates in speech recognition and achieving expert-level performance in medical image interpretation. There has been much excitement about the power of neural-network-based generative AI systems to produce language and imagery and to perform problem-solving. Applications of the technology are wide-ranging and include creating art, writing code, designing products, and helping with scientific hypothesis generation and research.

ML tools and technologies, including discriminative ML and foundational models and generative AI, offer distinctive benefits in health and health care applications. For example, the algorithm-driven technologies have demonstrated power and robustness for high-stakes diagnostic reasoning and therapy planning, and the generative AI technologies are showing early signs of potential for assisting with diagnostic and therapeutic reasoning and perhaps with even more prowess with natural language, illustrated by their strength in supporting tasks such as summarization of notes and report generation. It is particularly important to note that, notwithstanding the excitement about the most recent technologies, algorithm-driven ML continues to demonstrate its value with its complementary powers and applications in medicine and is also at an early stage of adoption.

IMPLICATIONS OF THE USE OF AI IN HEALTH, HEALTH CARE, AND BIOMEDICAL SCIENCE

The use of AI technologies that mimic or exceed human capacities is expected to be transformational in health care, along with almost every other sector of society. AI is being or may be used to support medical research, design new therapies, diagnose illness, identify personalized treatment plans, write patient care summaries, translate clinical advice for patient education, submit claims, appeal insurance denials, among a myriad of possibilities. Amid these advances of AI technologies and their applications in health, health care, and biomedical science, the criticality of human primacy, agency, and connection must be paramount. See Figure 2-2.

First, until there is adequate evidence that AI is equivalent or superior to human decision making within the context of each use case, and this performance can be safely and equitably sustained, AI systems may best be designed to support human decision making, as assisted by AI rather than ceded to it. Applying

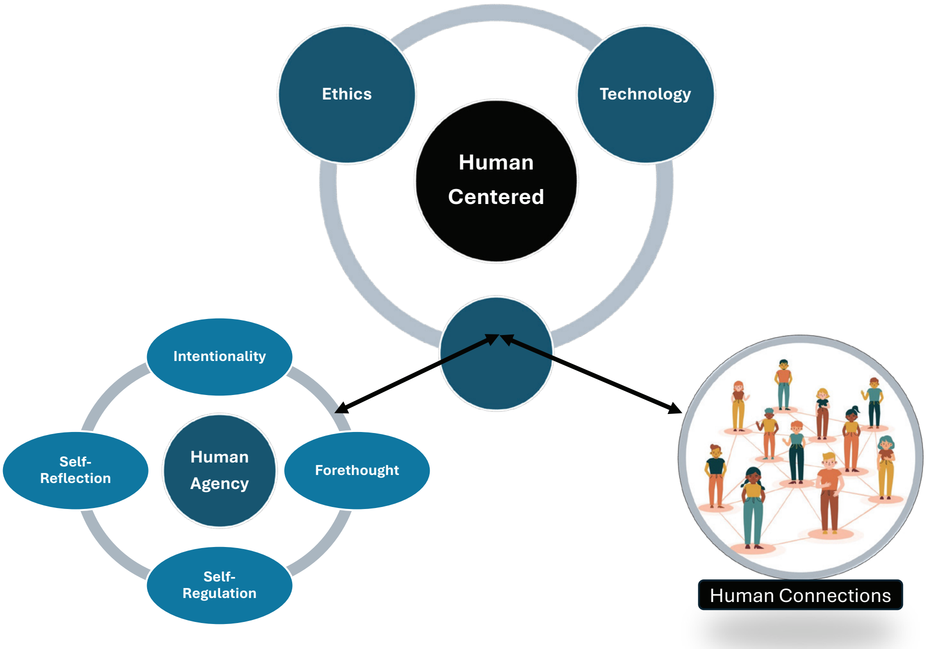

FIGURE 2-2 | Conceptualization of human centeredness emphasizing human agency and connections.
SOURCE: Bandura, 2006.

Human Factors Engineering principles and the concept of the copilot in aviation, AI developers can design systems that are subject to human oversight and communicate effectively with the human in the loop, yet still have assigned autonomous functions and clearly complementary skills and serve as a backup for human decision making (Sellen and Horvitz, 2024). Note, however, that humans can experience reliance bias from AI recommendations, and careful assessment of performance and feedback loops regarding AI-assisted human decision making in the context of use, and additional research into how to most effectively support human cognition with AI, is warranted (Jabbour et al., 2023).

Second, agency—defined here as the ability of individuals to make informed choices and take actions based on personal goals and values—is embedded in the LHS Shared Commitment, "Health and health care that is ENGAGED" (McGinnis et al., 2024) and is essential for AI systems developers to recognize and incorporate. Issues of personal control include both employing personal data in AI models as well as controlling recommendations or decisions made by AI systems (Li et al., 2022). Effective decision making is dependent on the understandability of AI-generated recommendations or decisions, including issues of risks, benefits, accuracy, reliability, and robustness of the underlying models, as well as the provision of well-calibrated uncertainties about inferences and recommendations to both patients and providers (Nori et al., 2023). Another important direction is the explanation of inferences and recommendations to physicians and patients, with appropriate references to information from the patient, the electronic health record, and the relevant literature (Sellen and Horvitz, 2024).

Finally, the maintenance of human connection is foundational to health (HHS, 2023b) and to trust in the healing relationship. Patients have a critical need for human connection when it comes to their health and important medical decisions, and physicians derive professional satisfaction and meaning in work based on these relationships (Hiefner et al., 2022). Under ideal practice conditions, clinicians have a broader contextual understanding of patients and their care goals than any automated tool. In addition, clinicians are ultimately accountable for health-related recommendations and therapies rendered to patients. Ideally, great AI enhances the human connection by both automating tasks and supporting human cognition through algorithms and naturalistic interfaces, ensuring that the human connection is more informed, meaningful, and decisive. Thus, when developing and deploying AI in health care settings, maintenance of "human connection" must be prioritized amid a world of rising automation.

HOW ADVANCED AI IS DIFFERENT THAN RULE-BASED DIGITAL HEALTH TECHNOLOGIES

Underlying technological differences in data processing, algorithmic characteristics, programming techniques, and computational resources have resulted in distinct differences in the capabilities of AI systems using ML or foundational models/generative AI (FM/GAI) systems (advanced AI) as compared to knowledge and rule-based (KRB) digital health technologies. ML-based AI is being developed and deployed in a rapidly evolving environment, with a new array of available tools, creating new opportunities and risks (Topol, 2019); and in the context of necessary attention to management of data and models, development and deployment, algorithms and infrastructure, ML-based AI has potential to scale in ways unimaginable with rule-based systems (Carnegie Mellon University Software Engineering Institute, 2021).

At a basic level, static, rule-based engines execute commands on structured data inputs, which are often limited in complexity, while discriminative ML and

FM/GAI systems rely on probabilistic statistical models and pattern recognition to predict designated outcomes using very large, growing, and changing datasets, which can include both structured data and unstructured natural language, images, audio, sensor, and video data. While all these systems may be run in extremely powerful, distributed cloud-computing environments, KRB digital health systems are less likely to be designed to take advantage of advanced computing power and distributed, disparate, and more complex data resources.

Unlike digital health technologies which require a human being to modify the rule-based engine, discriminative ML and FM/GAI systems may learn from available data, changing over time, without human intervention, creating challenges with explainability and transparency (Band et al., 2023) and requiring ongoing monitoring (Feng et al., 2022). Generative AI can create new information based on predictions gleaned from data, both historical and real-time and sometimes producing inaccurate results or hallucinations (Howell et al., 2024). Additionally, some AI systems have the capacity to interact with users, delivering human-like responses (Fernandes and Goldim, 2024), while others, like the AI systems behind self-driving cars, are capable of autonomous decision making (Bitterman et al., 2020). A side-by-side comparison of these differences in and implications of AI system features is outlined in Table 2-1.

These distinctive capabilities of various types of AI systems have resulted in a major shift in scale and scope of available systems, with a large and growing set of new tools being used to solve a wide variety of potentially novel challenges for a broad audience in the health arena, including clinicians, patients, researchers, developers, administrators, device manufacturers, and policy makers (AHA, 2019). The speed of new development and uptake of these tools has been rapid; in a recent survey commissioned by Microsoft, more than three-quarters of health care organizations reported using AI technologies (IDC, 2023). Open AI's chatbot, ChatGPT, had more than 180 million users with more than 1.6 billion visits per month as of February 2024 (Mortensen, 2024) many of whom may be turning to ChatGPT over traditional search engines for medical advice (Sandmann et al., 2024), thus making available the power of AI to a large majority of providers and patients in the United States.

Additionally, though not a systems feature, governance structures currently represent a point of divergence between rule-based digital health systems and payor AI systems. Rule-based systems are more mature and more limited in scope and have been carefully considered for inclusion and updating of existing governance frameworks, at the local, state, and national levels. AI governance structures can borrow from established risk management and health information technology governance frameworks, but some of the distinctive challenges in

TABLE 2-1 | Comparative Features and Implications of Various Categories of AI

System Feature	Knowledge and Rule-Based Systems	Discriminative Machine Learning	Foundational Models and Generative AI	Implications
Type of Tasks	Static predefined tasks based on predefined logic	Deliver predefined tasks based on training dataset(s). Require retraining for new tasks or application to new data	Capable of novel tasks based on extremely large training dataset. Retraining not required	Discriminative ML and FM/GAI offer opportunities to solve much more complex tasks.
Type of Data Employed	Typically, use structured or coded data and are often limited in complexity and number of sources of data As complexity and number of data features are more limited, data quality and characteristics (missingness, etc.) are easier to assess	Can employ structured and unstructured data, including text, images, genomic, audio, video, and multimodal sources. Can access data across multiple platforms Large numbers of data features create challenges in assessing data quality, missingness, variability, and availability across the full set of source data		Discriminative ML and FM/GAI systems may increase risk of exposure of personal health data, including protected health information. All technology types may introduce bias based on data characteristics or on how the features included in the logic or model are selected. , discriminative ML and FM/GAI systems are much more susceptible to these effects.

continued

TABLE 2-1 | Continued

System Feature	Knowledge and Rule-Based Systems	Discriminative Machine Learning	Foundational Models and Generative AI	Implications
Data Processing, System Learning, and Adaptation	Operate based on static, pre-defined rules and protocols. These systems do not adapt or improve unless manually updated and they are incapable of learning Generally curated to prioritize administratively or clinically meaningful features, which can limit performance but also improves interpretability and may limit overfitting to the local data context Challenge to keep these systems current and to harmonize content across rules	Employ data-driven, adaptable computational models to learn from vast datasets, including patient records, imaging data, video, audio, and genomic information, allowing AI to continuously update its predictive outputs		Discriminative ML and FM/GAI systems are based on complex data and adaptable over time. Some methods do not have explainable underlying logic (black box), and transparency and interpretability can be challenges. Because discriminative ML and FM/GAI systems and data change over time, resulting accuracy may change (and degrade) over time (model drift). KRB technologies are unable to learn and adapt to accumulating data.

TABLE 2-1 | Continued

<table>
<tr><th>System Feature</th><th>Knowledge and Rule-Based Systems</th><th>Discriminative Machine Learning</th><th>Foundational Models and Generative AI</th><th>Implications</th></tr>
<tr><td>Pattern Recognition</td><td>Require human interpretation and coding of pre-defined patterns

Rely on clinically or administratively meaningful data features that are human encoded and likely to be causal or associated with the purpose of the technology, improving interpretability but limiting performance</td><td colspan="2">Excel at identifying patterns in complex medical data, such as detecting anomalies in medical images, predicting disease outbreaks, or recognizing trends in patient symptoms</td><td>Discriminative ML and FM/GAI systems can be capable of making connections that are not readily transparent to end users, potentially leading to earlier and more accurate inference and application. However, these inferences may be subject to overfitting and lose accuracy outside of the population in which they were developed.

KRB technologies may miss informative data features that are non-intuitive to human developers.</td></tr>
<tr><td rowspan="2">Prediction</td><td rowspan="2">Wide use of historical risk calculators and rule-based tools

Can introduce bias and experience performance changes over time as data and clinical practice change</td><td colspan="2">Can predict patient outcomes based on comprehensive data analysis</td><td rowspan="2">Discriminative ML and FM/GAI systems may improve outcomes by forecasting disease progression, personalizing treatment plans, and identifying patients at high risk for complications.

Different challenges arise for classification versus accurate individual risk prediction.</td></tr>
<tr><td>Can present challenges with portability and adaptation to the local context as AI overfits to the source data context and requires evaluation, monitoring, and local adaptation</td><td>Has an improved capacity to include context in predictions</td></tr>
</table>

continued

TABLE 2-1 | Continued

<table>
<tr><th>System Feature</th><th>Knowledge and Rule-Based Systems</th><th>Discriminative Machine Learning</th><th>Foundational Models and Generative AI</th><th>Implications</th></tr>
<tr><td rowspan="2">Reliability and Robustness of Results</td><td rowspan="2">Consistently produce the same results with consistent inputs.

Unexpected inputs may result in clear error messaging as the inputs can be scored and tracked</td><td>Provides consistent results from consistent inputs</td><td>May yield inconsistent and inaccurate results (hallucinations) with consistent inputs</td><td rowspan="2">Discriminative ML and FM/GAI may produce inaccurate results that are difficult to discern from accurate results.</td></tr>
<tr><td colspan="2">Present significant challenges in determining etiology of change when the AI algorithm is a black box, and multiple source features are changing over time</td></tr>
<tr><td>Automation of Complex Tasks</td><td>Capable of automation of straightforward tasks, including scheduling appointments, generating billing, and basic data entry, but lack the sophistication to handle more complex administrative and clinical tasks</td><td colspan="2">Can automate complex tasks, including interpreting such things as radiology images and pathology samples, processing natural language in EHRs, and personalizing patient care plans</td><td>Discriminative ML and FM/GAI systems are better suited to model tasks as they become more complex. All types are susceptible to changes in the sequence and character of complex task processes and require maintenance and surveillance for intended function.</td></tr>
<tr><td>Interactivity and Human-like Capabilities</td><td>Offer static user interfaces without the ability to understand natural language or engage in interactive, human-like communication</td><td>Produce relatively static output formats and fixed integration into user workflows. Fixed inputs although the development of the inputs can be done in a fully automated and data-generated fashion</td><td>Can provide context-aware interaction with end users. May power interactive technologies such as data visualization or synthesis as well as virtual health assistants and chatbots, which can understand and respond to end-user queries</td><td>FM/GAI systems may be anthropomorphized, which may be a benefit or a harm, depending on the use case and the level of user sophistication.

Increasing end-user flexibility of interaction may come with unintended outcomes, both benefits and harms.</td></tr>
</table>

TABLE 2-1 | Continued

System Feature	Knowledge and Rule-Based Systems	Discriminative Machine Learning	Foundational Models and Generative AI	Implications
Scalability and Flexibility	Require significant modifications to adapt to new applications or scale with increased data. They are often less flexible and more challenging to customize	With the appropriate guardrails, may be scalable across clinical settings and can be tailored to specific clinical needs and patient populations (Pouyan, 2024) Lower barriers to customization		Discriminative ML and FM/GAI systems have shown that general-purpose systems can be developed at scale. However, overfitting to the development data environment and limitations in transportability of models require adaptation and customization for a local context of use (Lasko et al., 2024). For specific use cases, highly tailored KRB technologies may still have superior accuracy or reliability when compared to discriminative ML and FM/GAI systems.
Real-time Processing and Insights	Can provide real-time data processing with lower computing costs due to simplicity of rule-based systems	Provides real-time data processing and actionable insights, can integrate dynamic, changing data elements and variable processing intervals over time Requires attention to scalable architecture to manage computing costs		All types of systems present challenges in ensuring that the data stream has sufficient fidelity and quality for the context of use due to real-time use.

continued

TABLE 2-1 | Continued

System Feature	Knowledge and Rule-Based Systems	Discriminative Machine Learning	Foundational Models and Generative AI	Implications
Testing and Ongoing Monitoring	Assessed based on the system's functional conformance to specifications; Identified errors and new features typically addressed in batch mode	Assessed based on statistical (quantitative) and qualitative model quality relative to requested task and desired outcomes	Assessment of outputs may be more complex due to changes to output over time Assessment relative to desired outcomes are essential	All systems require ongoing monitoring, but discriminative ML and FM/GAI systems have more complex requirements to ensure safe, equitable performance. Ensemble AI automation (e.g., use of one model to monitor another model) can be applied to support monitoring and performance maintenance of complex AI systems. AI and economic performance measures.

NOTE: EHR = electronic health record; FM/GAI = foundational models/generative AI; KRB = knowledge and rule-based; ML = machine learning.

these technologies have yet to result in mature AI-adapted governance structures. In addition, while access to large AI systems may be limited to organizations with significant resources, creating potential equity issues, access to AI systems has been democratized because of the availability on cell phones and via the Internet, making adherence to best practices more difficult to detect and address in some settings (e.g., physicians using LLM tools to summarize notes or to appeal insurance denials).

RISKS ASSOCIATED WITH THE USE OF AI

Novel attributes and rapidly changing capabilities of AI systems relative to rule-based digital health technology yield both impressive opportunities for new functionality along with both overlapping and risks (FDA, 2021, 2022). As briefly outlined in Table 2-1, there are considerations applicable to rule-based, discriminative ML, and FM/GAI-based systems which include data privacy and security; introduction of a variety of forms of bias from the process of tool development; the intended function, and the context of use; reliability of results;

and impacts of automation on human experience and the workforce. Other considerations are more specific to discriminative ML, and FM/GAI-based systems, such as explainability and transparency; anticipated and unexpected bias from the training source data; anthropomorphizing technology, including deep fakes; model drift and the need for ongoing monitoring; and real-time processing and scalability. All these challenges have the potential to create snowball-like impacts for any of the issues noted here and negatively impact equitable access.

Data Privacy and Security

Massive amounts of data are essential for advancing AI in health and health care (Mandl et al., 2024) as they enable the development of sophisticated models capable of delivering personalized treatments, predicting patient outcomes, developing new therapies, and improving efficiency. However, the aggregation, management, and licensing of these extensive datasets introduce significant privacy and cybersecurity risks. Safeguarding patient data requires robust encryption, strict access controls, and adherence to regulatory standards to prevent unauthorized access and data breaches. Poorly understood, broad consents for use of personal data that consumers have granted to big tech firms in exchange for "free services" (e.g., chatbots, smart watches, and so forth) also present risks to individual privacy, yet remain unregulated. The licensing and sharing of health data by health delivery systems must balance the benefits of data-driven innovation with the ethical obligation to protect patient confidentiality and agency. For example, collecting deeply personal information—voice, eye movement, facial expressions, body movement, and reaction times—and employing these data in behavioral health applications hold promise to significantly improve access to treatment as well as quality of care, but also presents substantial risks for harm if these data are accessed and used nefariously (Olawade et al., 2024).

Bias

AI systems reflect the structural biases embedded in the data on which they are trained as well as the conscious and unconscious biases of developers and end users of AI. If the training data are biased or unrepresentative of the target population, including the predicted outcome of interest, AI systems can perpetuate and even exacerbate existing health disparities (Obermeyer et al., 2019); and the models may perform poorly on groups under-represented in the data. For example, facial recognition technologies developed on non-diverse populations, perform poorly on non-White populations (Grother et al., 2019). If used in diagnostic algorithms (Qiang et al., 2022; Wu et al., 2021) without eliminating the bias, it could lead to misdiagnoses and inappropriate treatments in health care applications. For example, lack of representative training data in pulse oximeters led to lower accuracy among patients with darker skin (Sjoding et al., 2020) and subsequently to delayed recognition of hypoxia during the COVID-19 pandemic (Fawzy et al., 2022). In health care, this can result in distorting model output, including recommendations or decisions, impacting patient care and outcomes.

Explainability and Transparency

As opposed to rule-based tools which were explicitly designed with explainability and interpretability in mind, many ML-based models, particularly those using DNNs, by nature behave as "black boxes," making it difficult, even for their developers, to understand how they arrive at specific determinations (Saeed and Omlin, 2023). It should be noted that there are opportunities to combine more "black box" methods with either explainable discriminative ML or rule-based systems, and this is an area of ongoing research. However, in some contexts of use, this lack of explainability can hinder transparency (sharing with impacted parties information about data sources, methodologies, and testing results) and thus negatively impact trust and acceptance among health care professionals and patients. Moreover, while naturalistic interfaces in generative AI tools may support chain-of-thought reasoning, the identification and correction of errors may still prove challenging.

Reliability of Results

Given the potential for AI dependence on training datasets with particular attributes, some AI systems may falter when data not conforming to the training dataset attributes are presented (Finlayson et al., 2021), yielding inaccurate results, a concept known as overfitting on the training data attributes. This is particularly problematic for health care diagnostic and treatment recommendations, where outcomes have potentially life-altering impacts. This challenge is exacerbated by the issues of explainability and transparency noted above (Saeed and Omlin, 2023). New AI techniques are being employed to reduce these risks, including the use of more robust validation datasets as well as application of generative models, federated learning, and synthetic datasets (Hong et al., 2023).

Impacts of Automation on Health Care Decision Making, Human Experience, and the Workforce

AI systems capable of autonomous decision making pose the risk of over-reliance on technology for clinical decision making, loss of human connection (Quinn et al., 2021) and significant changes for the health workforce, including retraining needs and job loss (Reddy, n.d.). Among AI-assisted tools that provide recommendations or cognitive support, there is also a risk of users depending excessively on AI recommendations or visualizations, resulting in bias and over-reliance (Jabbour et al., 2023), as well as the potential for de-skilling (Aquino et al., 2023). And, while in some situations, interaction with AI chatbots could improve equity of access (Habicht et al., 2024), it is possible that when engaging with chatbots rather than humans, patients may feel isolated and their trust in the healing relationship could be eroded.

As AI systems become increasingly adept at performing tasks such as diagnostics, treatment planning, and administrative workflows, the traditional roles and skill requirements within the health care workforce are set to undergo profound changes (Davenport and Kalakota, 2019). Initially, AI is poised to ameliorate the workload of documentation, theoretically allowing clinicians to more fully engage with their patients (Tierney et al., 2024). Furthermore, the integration of AI could lead to important shifts in workforce composition. Various members of the health care workforce, supported by advanced AI tools, might take on responsibilities traditionally held by doctors, thereby altering the professional landscape. While AI may offer solutions for workforce shortages and enhance efficiency of operations, job displacement concerns as well as issues associated with electronic health record (EHR) use may dampen the appetite for AI adoption among physicians. There is also a need for training in use of AI, continuous education, and retraining. Clearly, workforce and educational implications are significant and are addressed in a later section in this publication.

Anthropomorphizing Technology

Anthropomorphism ascribes human qualities—those not seen in traditional systems including moral character, status, and judgment quality—to AI systems that do not in fact possess them (Placani, 2024). Yet, some AI outputs are presented with human-like characteristics such as emotion, physical appearance, and even self-consciousness (Steerling et al., 2023) and can be offered up in a conversational manner that mimics human communication, potentially leading to emotional response or overconfidence in outputs by the end users. While limited research has been done in this arena (Liu and Tao, 2022), over-reliance or declining trust are both potential downstream risks that must be considered. In addition, increasingly granular data collection from facial expressions, voice patterns, and other human behavior, could be used to create AI outputs that are virtually indistinguishable from reality, so-called "deep fakes," that can be used to impersonate an individual and could result in serious personal harm (Mirsky and Lee, 2021). Navigating these challenges is critical to harnessing the full potential of AI in health care while maintaining public trust and ensuring patient safety.

Data and Model Drift and Ongoing Monitoring

AI models that continuously learn, adapt, and encounter new data over time can improve their performance but must be properly monitored. Poor performance can result in suboptimal care, misappropriation of resources, and safety risks (Davis et al., 2017). Changes in data patterns or the emergence of new medical knowledge can render existing models outdated, not helpful, or even harmful (Kore et al., 2024). Conversely, online AI or ML that learns continuously from incoming data presents a different risk of drift in which the performance of the tool and outputs could shift in response to the data but not the underlying design or health goal. In all cases, implementing robust monitoring and updating mechanisms are essential to ensure that AI systems remain accurate and effective (Davis et al., 2022).

Real-Time Processing and Scalability

Given the real-time processing capacity of AI within clinical systems (e.g., colonoscopies and surgery) (Topol, 2019), any issues with the model (e.g., drift) may not be detected before clinical decisions are made, potentially leading to patient harm. Scalability presents an escalation of risks mentioned above, including data privacy, bias, model reliability, and model drift. Scaling AI also involves increased resource utilization. The use of AI in health care requires substantial energy consumption and affects the carbon footprint associated with training and operating advanced ML models (Jia et al., 2023). As health care (and other aspects of society) become more dependent on AI, the potential for increasing the impact of AI on the climate cannot be ignored. Data centers housing AI systems require vast amounts of electricity for computation and cooling, contributing to greenhouse gas emissions. This results in a careful tradeoff of building new AI resources versus retraining or adapting existing resources given the computational and environmental impact of large model development.

Equitable Access

Substantial organizational resources are required to develop, acquire, implement, and govern health AI responsibly. Beyond the cost of development or procurement, needs for technical, analytic, and compliance expertise are typically expensive and can be out of reach for small, rural, or poorly resourced health organizations. One study demonstrated that AI was more likely to be incorporated into medical care in higher-income, metropolitan areas (by zip code) with academic medical centers (Wu et al., 2023). As a result, there is a risk of inequitable access to potentially life-extending AI tools for patients seeking care at lower-resourced organizations, further exacerbating existing inequities in care.

New and Emerging Tools, Including Large Language Models

"The foundation model (FM) is a family of machine artificial intelligence (AI) models that are generally trained by self-supervised learning using a large volume of unannotated dataset and can be adapted to various downstream tasks" (Jung, 2023). Foundational LLMs are distinguished by using a set of AI technologies and developing meta-parameters and internal representations entirely from a set of source data. Foundational LLMs can be adapted through few-shot learning (an ML technique that trains AI models to make predictions using a small number of labeled examples) and reinforcement learning for specific use cases. LLMs present new classes of risks not previously encountered in software as medical devices (FDA, 2021, 2022) and relying on them for high-risk health care functions will require special testing, monitoring, and oversight. Foundational LLMs can hallucinate by providing inconsistent results or omissions, and the quality of their responses can fluctuate over time (Perković et al., 2024).

The largest foundation language models also present risks for introducing or perpetuating bias. They have been trained on vast datasets, many from the public Internet, that reflect societal discourse replete with cultural, political, and scientific biases (Zack et al., 2024). Additionally, LLM development can employ an alignment approach called reinforcement learning with human feedback that involves human trainers providing feedback on the model's outputs. Human trainers may unconsciously introduce biases during this process or be instructed in a manner that ultimately shapes model responses.

Yet, for consumers and patients, not since Google Search (Tang and Ng, 2006) has there been a widely available information technology innovation as impactful for health as LLM-based chatbots. They may fill in key gaps in consumer access to expert health advice. However, they introduce risks that warrant careful consideration. Privacy and data security are paramount concerns, as chatbots often collect, store, and use sensitive personal information, perhaps in a manner not apparent to the user. Miscommunication and errors can arise from their limited understanding of context and nuance, potentially leading to misinformation or inappropriate responses. Additionally, consumer and patient reliance on chatbots for critical tasks could reduce important medical oversight. Ensuring ethical use, robust security measures, and continuous improvement in their accuracy and fairness is essential to mitigate these risks.

Regulation

Keeping pace with innovation poses a significant challenge for regulators; technologies, such as generative AI tools like ChatGPT, which produce unvalidated outputs, provide a clear example of a tool that may have unpredictable impacts on health care systems (Bouderhem, 2024).

Balancing innovation with transparency of the AI inputs, outputs, and expected operation as well as transparency of disclosure of AI use, equity, and safety needs are daunting, and much work is being done by AI researchers to address the data and technical challenges; simultaneously, governance efforts—local, national, and international, and including this AI Code of Conduct (AICC) framework—are being developed to address the risks to ensure that the benefits are realized evenly across society (HHS, 2024a; Hong et al., 2023; Quinn et al., 2021).

ALIGNING AI GOVERNANCE EFFORTS FROM THE ORGANIZATION TO THE GLOBAL ARENA

AI governance has been defined as "a system of rules, practices, processes, and technological tools that are employed to ensure an organization's use of AI technologies aligns with the organization's strategies, objectives, and values; fulfills legal requirements; and meets principles of ethical AI followed by the organization" (Mäntymäki et al., 2022). Governance can be employed at an organizational, local, national, international, or global level. The continuum of the forms of governance, from most conceptual to most enforceable and based on the work of Mills and colleagues (2023), is outlined in Figure 2-3.

Given the potentially transformative benefits of AI to improve health, health care, and biomedical science, along with the risks of AI outlined above, a central goal of the National Academy of Medicine AICC project is the alignment on a set of foundational principles and commitments designed to promote responsible use of AI across the health sector in the United States. There is a broad recognition and

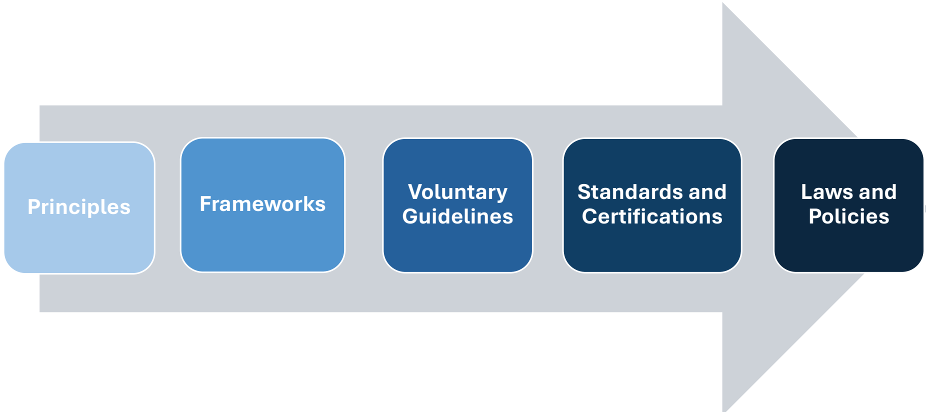

FIGURE 2-3 | Continuum of the forms of governance from most conceptual to most enforceable.
SOURCE: Conceptually adapted from Mills et al., 2023.

clarity among the authors that while ethical principles and commitments provide a governance starting point, organizations that develop and/or deploy AI in the health sector will also require more detailed guidance enabled through broader accountability frameworks, standards, and policies. Additionally, AI systems and their accompanying opportunities and threats are not bounded by local, state, or national borders. Therefore, these alignment efforts across the governance continuum must also be viewed from a global perspective.

Several factors are driving and influencing global governance of AI. AI development involves transnational actors, particularly multinational corporations (Kshetri, 2024), for whom common rules could ease regulatory burden (Tallberg et al., 2023). AI has the potential to provide significant economic advantages, leading to competition in the international race for AI dominance, potentially stifling cooperation (Vijayakumar, 2023). AI systems create externalities that transcend national borders, necessitating international cooperation for effective regulation to create a level playing field in the interest of all parties (Tallberg et al., 2023). In addition, AI is financed, developed, used, and governed by a diverse set of international actors with varying resources and sometimes disparate or competing values, incentives, and motivations, creating a complex governance landscape (Gianni et al., 2022).

While there are growing efforts to promote international collaboration, AI governance efforts to date have been largely nationally focused. For example, the Department of Health and Human Services (HHS) Office for Civil Rights (OCR) and the Centers for Medicare & Medicaid Services issued a final rule under the nondiscrimination section of the Affordable Care Act (HHS, 2024a), specifying that patient-care decision support tools, including ML and AI models, must not exhibit "discrimination based on race, color, national origin, sex, age, disability" (HHS, 2024a). Discriminatory practices are prohibited and will be enforced under existing federal laws, including Title VI of the Civil Rights Act of 1964 (Department of Justice, n.d.). However, the specific policies and procedures needed to comply with the statute still require substantial development. Further complicating the ecosystem in the United States, state-level regulations are being enacted or considered. Across 40 states, a broad range of bills to address the risks of AI were introduced in 2024, and resolutions were adopted or bills passed in six states, along with Puerto Rico and the U.S. Virgin Islands (National Conference of State Legislatures, 2024).

The Organisation for Economic Co-operation and Development (OECD) AI Policy Observatory recorded more than 700 national AI governance initiatives from 60 countries and territories (Tallberg et al., 2023). Despite having developed national AI strategies, many countries are now struggling to establish more robust governance, which require careful consideration of issues including risk-based frameworks, licensing agreements, liability structures, standards, and research and development support (World Economic Forum, 2024).

AI presents some challenges and special considerations for governance.

1. **Uncertainty:** The performance of novel applications of AI techniques and tools is difficult to predict, including the intended and unintended outcomes of such applications. This presents challenges in proactive governing and confidently assessing the best approach to governance. Considerations include the types of governance models that are most appropriate, which aspects of AI should be included or excluded in governance, and at what level governance should occur—regional, national, or international level (Sepasspour, 2023).
2. **Rapid evolution:** Breakthroughs in AI development have and are expected to continue to outpace the capacity of traditional governance approaches, which typically lag new technologies. This is in part because of the lengthy legislative process and in part due to the inability of regulators to assess rapidly the benefits and risks associated with the technology (World Economic Forum, 2023). Adaptation and collaboration with the private sector will be required to respond to this environment (Best et al., 2024).
3. **Societal impact:** Given the apparent "general-purpose" nature of AI (Crafts, 2021), the scope of impact on society is quite broad and will be felt across most industries and sectors. The historical approach to technological innovation is one primarily focused on profit, which can result in societal

harm; consideration for alignment of AI investment with the public interests—"economic, societal, and national security objectives"—is warranted (Mazzucato et al., 2022).

4. **Competition:** Significant private investment in AI development (Ahmed et al., 2023b) can create fierce competition for market advantage, and result in a focus on speed over safety or profit over ethics (Dafoe, 2018). Additionally, such competition can create "winner-take-all" or "winner-take-most" scenarios, concentrating benefits, both economic and political, in a small number of organizations or countries (Muro and Liu, 2021).
5. **Fragmented governance:** Various non-governmental parities are engaged in the creation and use of AI technologies and tools, and many are actively working on AI standards and governance frameworks. Broad collaboration and inclusive participation of parties with international influence and/or their deep technical expertise will be essential in the ongoing development of AI (Schmitt, 2022).
6. **Geopolitical competition:** Competition is not limited to commercial interests; international competition is focused on geopolitical wealth and power advantages (Dafoe, 2018). In such a high-stakes environment, where national security is deemed a risk, cooperation to establish AI governance standards and regulatory frameworks can prove challenging, particularly among nations that are not ideologically aligned, as is seen in the oppositional positioning on AI between the United States and China (Roberts et al., 2024).
7. **Lack of consensus on priorities:** Key stakeholders do not agree about which challenges associated with AI should be prioritized or about the policies that would address the issues. For example, the United Kingdom is focusing on existential harms, while the European Union is emphasizing near-term risks such as bias (Roberts et al., 2024).

In response to these challenges, a growing number of public and private international organizations have engaged in governance efforts from convening experts to developing international regulation to fill the gaps. As documented in the draft AICC literature review, there is significant synergy around the underlying principles of responsible AI (Adams et al., 2024). However, numerous approaches to governance are emerging. For example, at the G20 summit in New Delhi in September 2023, a human-centric AI governance framework was proposed by the Indian Prime Minister, and an AI risk monitoring body based on the Intergovernmental Panel on Climate Change was suggested by the president of the European Commission (Kshetri, 2024). Table 2-2 provides a sampling of key international AI governance efforts across the continuum since 2019, including

TABLE 2-2 | A Sampling of Key International AI Governance Efforts Since 2019

Form of Governance	Year	Sponsor	Description	Link
Principles	2019	European Commission	High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence	https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
	2019, Updated 2024	Organisation for Economic Co-operation and Development (OECD)	Adopted by OECD member countries focusing on ethical AI. G20 leaders committed to these principles	https://oecd.ai/en/ai-principles
	2021	United Nations (UN) Educational, Scientific and Cultural Organization	Adopted by all 193 member states to guide legal frameworks for AI ethics	https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence
	2021	World Health Organization (WHO)	Expert report and recommendations for principles to ensure AI works to the public benefit of all countries	https://www.who.int/publications/i/item/9789240029200
	2023	G7, OECD	Presentation of background and principles for governing generative AI for G7 leaders in Japan	https://www.oecd-ilibrary.org/docserver/bf3c0c60-en.pdf
	2023	United Kingdom	Bletchley Declaration, signed by 28 countries during the UK International Summit for AI Safety, emphasizing the need for collective management of AI risks	https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023
	2023	United Nations	Interim Report: Governing AI for Humanity	https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf
	2024	WHO	AI guidance on Large Language Models	https://www.who.int/news/item/18-01-2024-who-releases-ai-ethics-and-governance-guidance-for-large-multi-modal-models

Framework	2023	Council of Europe	Draft framework on AI and human rights, democracy and rule of law	https://rm.coe.int/cai-2023-28-draft-framework-convention/1680ade043
Regulation	2024	European Union	First international regulation on AI	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL_202401689
Multilateral Initiatives	2020	Global Partnership on AI	Launched by 15 countries to support ethical AI adoption	https://gpai.ai/about
	2021	EU-US Trade and Technology Council	Formed to coordinate activities in AI and other technologies	https://digital-strategy.ec.europa.eu/en/library/ai-public-good-eu-us-research-alliance-ai-public-good
	2023	BRICS AI Study Group	Formed to study global equity in AI	https://www.reuters.com/world/chinas-xi-calls-accelerated-brics-expansion-2023-08-23
	2023	UN	Formed to support international collaboration on AI governance	https://www.un.org/sites/un2.un.org/files/231025_press-release-aiab.pdf

the work of OECD, the United Kingdom, the European Union, the Council of Europe, the United Nations, and the World Health Organization.

Despite a lack of global consensus on the best approach, there is a clear need and momentum for a governance framework for AI. Taking definitive action, the most stringent international governance of AI is being implemented in Europe via the AI Act, which was passed in March 2024 to ensure that AI development, deployment, and use in the European Union promotes innovation and EU values, while mitigating the risks of AI (Artificial Intelligence Act, EU Regulation 2024/1689). Furthermore, in April 2024, the European Union and the United States agreed on a risk-based approach to AI governance to ensure "safe and trustworthy AI" produced by the United States and the European Union (European Commission, 2024a).

AI governance involves a set of interwoven issues that require coordinated regulation and a nuanced policy agenda. Overarchingly, more international collaboration would be beneficial. This agenda should consider the promotion of innovation to address complex issues and prevent harmful proliferation in the context of a competitive marketplace. Given that the private sector currently controls most aspects of AI, its governance is particularly complex and politically sensitive. AI governance on the global level could be complicated by the dominance of major tech firms or the growing involvement of smaller players, such as emerging technology companies (Kshetri, 2024). In part to promote trust in AI tools and technologies, a variety of stakeholders, including big tech, researchers, and non-governmental organizations have taken steps to define good governance and use those work products to inform public policy. For example, the Partnership on AI (PAI) was formed in 2016 to develop guidance and inform public policy (Schmitt, 2022), and in 2023, the Frontier Model Forum[1] was established to ensure "the safe development and deployment" of cutting-edge AI technologies, known as frontier models. However, the complexity of self-governance by technology companies—particularly in the face of misaligned incentives—was demonstrated by resignation from PAI by Access Now, a digital rights organization, as they "did not find that PAI influenced or changed the attitude of member companies or encouraged them to respond to or consult with civil society on a systematic basis" (Access Now, 2020).

Multiple global bodies and coalitions have begun to formulate an approach to governing AI, with a primary focus on principles and frameworks. While rules and standards for the development and use of AI are being established, keeping pace with AI's potentially disruptive effects will require major advancement in governance approaches (Sepasspour, 2023). Governing AI effectively will require

---

[1] See https://www.frontiermodelforum.org/about-us (accessed April 4, 2025).

rapid cycle learning, adaptability, and correction, just as AI has the capacity to rapidly evolve and autonomously self-improve (Bremmer and Suleyman, 2023). The history of Internet governance provides an example, elucidating the potential and the constraints of governance innovation driven by societal and technical advancements while simultaneously emphasizing the critical role of the multistakeholder approach embraced by both national governance bodies and Internet Corporation for Assigned Names and Numbers (Almeida et al., 2023). Applying these lessons and focusing on alignment will support the advancement of AI governance, locally, nationally, internationally, and ultimately globally.

APPLYING LESSONS FROM EHRs' ADOPTION TO THAT OF AI

In addition to heeding lessons about governance, careful consideration of the experiences of broad adoption of technology into the health care workflow is warranted. In recent years, AI tools in health care, like EHRs preceding them, have become available commercially and have been met with both hope and skepticism (Cary et al., 2023; Lindsell et al., 2020; Matheny et al., 2020).

Similarly, early discussions about EHRs generated unwarranted claims about benefits (Bates and Gawande, 2003; Hillestad et al., 2005; Lohr, 2005)—consistent with the Gartner Hype Cycle, a widely applied management framework for the consideration of new innovations (Gartner, n.d.). This framework suggests that new innovations follow a replicable pattern whereby an innovation trigger yields inflated expectations followed by disillusionment, improvement, and eventually productivity and broad adoption. EHRs have certainly been the focus of high expectations and disappointments among end users. In this context, comparing EHRs to an unrealistic ideal rather than viewing them as a tool capable of contributing to incremental progress often fueled inflated expectations and subsequent frustration. For example, while a small study of early EHR adopters demonstrated that 81% of clinicians found the EHR to be superior to paper records (Kaelber et al., 2005), the 2021 National Electronic Health Records Survey (NEHRS) estimated that only 25% of physicians were very satisfied and 35% somewhat satisfied with their EHR system (CDC, 2021). Although the journey to an ideal EHR is far from over and similar risks of comparisons to an idealized target exist with the deployment of health AI, considering key lessons that led to broad EHR adoption by U.S. physicians is a valuable exercise as health AI adoption is burgeoning. These lessons are generally organized into two categories: (1) understanding the rationale for adopting health care AI, and (2) translating behavioral intent into practice.

Understanding the Rationale for Adopting Health Care AI

The Theory of Planned Behavior (TPB) (Ajzen, 1985) provides a useful scaffolding for understanding technology adoption and may be a practical construct for scaling health AI use. In the context of EHR adoption, TPB addresses (1) attitudes toward EHRs based on available evidence; (2) establishing an expectation to support adoption (subjective norms); and (3) perceived behavioral control—the belief that a behavior is under the control of the individual—which is influenced by an environment that fosters behavioral control and behavioral intention.

Using Evidence to Impact the National Dialogue

Although early clinical information systems were developed in the 1960s (Atherton, 2011), development and broad adoption was limited for decades. According to data provided by the Office of the National Coordinator for Health IT (ONC), in 2008 fewer than 20% of hospitals and physicians had adopted EHRs (ONC, n.d.). The slow EHR adoption trend to that point was due, in part, to the lack of a clear rationale and supporting evidence for adopting an EHR in settings outside large hospitals. A series of Institute of Medicine (IOM, now the National Academy of Medicine [NAM]) reports that focused on health care quality were among the most influential compendia recommending the adoption of EHRs to minimize errors associated with missing data, erroneous data, unavailable clinical guidelines, and siloed patient information (IOM, 1991, 2000, 2001, 2003, 2007). These reports summarized the science around medical errors and poor health care quality. Following these reports, there was a robust conversation in the lay press and the scientific literature to better understand the need for funding (Anderson, 2007) and the potential benefits, pitfalls, and unintended consequences of the use of EHRs (Ash, 2004; Berg, 2001; Han et al., 2005). Importantly, literature from the major hospital systems with internally built EHRs described the efficiency and safety benefits, as well as challenges with the technology, shifting workload, and data security and privacy concerns (Chaudhry et al., 2006). Additionally, the National Research Council published a report in 2009 that identified "persistent problems involving medical errors and ineffective treatment.... Many of these problems are the consequence of poor information and technology (IT) capabilities, and most importantly, the lack of cognitive IT support" (NRC, 2009). Federal legislation and regulation followed; The Health Information Technology for Economic and Clinical Health (HITECH) Act (HHS, 2009) provided financial incentives to support the cost of EHR adoption and encourage adoption and meaningful use of EHR technology with penalties for those that failed to do so (Blumenthal and Tavenner, 2010).

With regard to clear rationale and evidence for use of AI in clinical settings, like with EHR adoption, there are similar questions and concerns about the validity of recommendations (Cary et al., 2023; Habib et al., 2021; Obermeyer et al., 2019), potential issues with trust and reliability (Esmaeilzadeh et al., 2021; Liu and Tao, 2022), job displacement (Howard, 2022; Petersson et al., 2022), and ethical infractions (Li et al., 2022; Sartori and Theodorou, 2022). While many experts are touting AI's benefit for disease prediction (Kumar et al., 2023) and equitable access to health/disease advice (Ayers et al., 2023; Kaur et al., 2024; Kurniawan et al., 2024), the most value in the near term may be in mitigating health care provider burnout (Borna et al., 2024; Levy et al., 2022; Liu et al., 2023; Sallam, 2023). Evidence regarding the potential benefits, risks, and costs of partnering with AI in clinical settings is needed. Costs should be described and assessed broadly, and include financial, time, and workflow reduction estimates, to make transparent the degree to which benefits outweigh the risks and costs. This assessment may be an important part of improving attitudes toward using AI in clinical settings, as with EHRs. Many concerns about adopting EHRs were overcome through scientific evidence and communication to various stakeholder groups, it may be advisable to do the same with health care stakeholders to surface and address concerns about AI in health care over time. In addition, evidence generation will create an evidence-based practice of AI in health and health care that may generate reimbursement for its use, which could further accelerate use, and even more importantly, democratization of use in low-resource environments.

Subjective Norms

The health care system is enmeshed in the Era of Entanglement (Johnson et al., 2021), whereby various stakeholder expectations dictate how technology is to function, and health care providers struggle to navigate an ecosystem riddled with regulatory, quality reporting, and social determinants requirements that impact care decisions and delivery. This complex landscape exacerbates the social pressures and expectations health care providers perceive surrounding AI adoption. Colleagues from one setting may de-prioritize concerns about biased data in favor of improved access to tools that minimize burnout, while institutional leadership may discourage using these same tools because of concerns about sharing health information with non-covered entities that may lack strong privacy policies. Although federal incentives were critical to establishing supports and expectations to advance widespread adoption of EHRs across the United States, establishing a culture of acceptance for EHR adoption has been difficult, and challenges with usability and satisfaction persist (Holmgren et al., 2024).

Concerns about the trustworthiness of AI will also need to be overcome. Wherever it has been studied, concerns about reliability, fairness, and bias continue to surface. This is true in lay press as well as academic publications (Angwin, 2023; Sanders, 2024), adding kindling to the combustible notion that AI is not ready for any use at scale, despite some emerging evidence in medical journals suggesting otherwise in specific contexts, such as radiology.

Perceived Behavioral Control

An important lesson from EHR adoption is that many providers have arguably not achieved self-efficacy (confidence in their ability to learn and use EHR systems), resource availability (access to materials that maintain self-efficacy and technological resilience), and a realistic recognition of obstacles, such as technical difficulties, potential disruption to established workflows, and human factors engineering being a work in progress. This perceived lack of behavioral control has contributed to issues such as burnout (Melnick et al., 2020; Moy et al., 2021). These consequences are often attributed to design issues among the vendor community; however, there are data suggesting that health care providers who complete EHR skills training may have improved efficiency, time, and lower rates of burnout versus those who have not completed that training (Lee et al., 2023; Robinson and Kersey, 2018). As AI is implemented in the health care workflow, it will be important to provide necessary resources to improve perceived behavioral control, including education, time to learn and incorporate the technology into the workflow, technical support, and ongoing financial support. Technical support for AI-augmented clinical tools (as currently conceived) will need to consider new failure modes, such as feedback (Aikens et al., 2024) or data biases (Obermeyer et al., 2019; Temple and Rowbottom, 2024).

Translating Behavioral Intent into Practice

Perceived behavioral control is necessary but not sufficient to ensure the widespread adoption of new technologies required for meaningful transformation. While health care providers may feel capable of implementing AI tools, actual adoption requires more than just perceived capability; true transformation demands a multifaceted approach beyond mere availability and perceived ease of use. Based on lessons learned from EHR adoption, successful behavioral change required private and public messaging and assistance investments. As with EHR adoption, successful implementation of health AI may also require regulatory pressure to create technology standards, incentivize national adoption, and even impose penalties for failing to adopt normative approaches to care using AI. Finally, long-term and sustainable success may benefit from viewing AI in health care as a Learning Health System (LHS) initiative, involving numerous stakeholders, standardized evidence generation, and summative feedback to developers, leading to data-driven innovation and improvement.

Promotion of Use Through Standards, Certification, and Technical Support

Additionally, promoting behavioral control, HITECH established a technical support infrastructure through regional extension centers (RECs). Their mission was to provide state and regional support, including expert knowledge about certified EHRs, adoption strategies for small group practices, and other nuanced help to ease providers into using EHRs in their offices. RECs appear to have added value and some degree of self-efficacy (Furukawa et al., 2015; Green et al., 2015; Lynch et al., 2014; Riddell et al., 2014).

Soon after the passing of the HITECH Act, HHS issued regulations adopting "an initial set of standards, implementation specifications and certification criteria" for EHRs and creating a voluntary certification program (ONC, 2010). These standards and criteria are routinely updated to adapt to changes in standards and policy priorities, including adding requirements for algorithm transparency in the most recent final rule (ONC, 2024a,b). The ONC Health IT Certification Program, which was linked to the meaningful-use EHR incentives programs adopted by the Centers for Medicare & Medicaid Services, continues today and remains crucial for assuring that EHRs incorporate required standards and functionality (ONC, 2011, 2016). This program manages an authorized testing laboratory and an ONC-authorized certification body responsible for issuing certifications and ongoing surveillance. The evidence for the impact of this program is mixed (Bowes, 2014; Pylypchuk and Johnson, 2022; Ratwani et al., 2024); however, as a result of this program, recent legislation has approved a voluntary EHR certification program for pediatric EHRs, suggesting that at least some communities find it to be of value (Thompson et al., 2022).

HITECH's Impact

The multipronged approach, designed to advance from behavioral intention to full-scale adoption succeeded in many ways, including increasing EHR adoption (Cohen, 2016; Washington et al., 2017), impacting inpatient mortality rates (Trout et al., 2022), impacting care quality and efficiency (HHS, 2014), and catalyzing new discoveries through programs such as PCORI and AllofUs that rely on voluminous EHR data. However, the rapid pace of EHR adoption has been associated with nursing and physician burnout (Halamka and Tripathi, 2017; Washington et al., 2017), leading to initiatives such as Getting Rid of Stupid Stuff

(Ashton, 2018), and ClickBusters (McCoy et al., 2022), as well as the American Medical Informatics Association (AMIA) 25×5 program to decrease the burden of health care provider documentation (Johnson et al., 2021; Moy et al., 2021). These initiatives were enabled by the ongoing evaluation and improvement science focusing on the EHR and will continue into the foreseeable future.

While HITECH defined standards and incentivized adoption, cognitive supports (including user design, workflow integration, and clinical decision support) and interoperability were not adequately considered. Follow-up legislation, the 21st Century Cures Act (2016), included directives to reduce administrative (documentation) burden, improve interoperability, reduce information blocking, and increase patient access to their medical records. Efforts to incorporate user design and provide cognitive support were encouraged but not explicitly called out for regulation.

AI and Translation into Practice

Although the underlying drivers (e.g., incentives, political and economic forces) between EHR and health AI adoption are significantly different, steps to implement AI in health care appear to be following much of the same blueprint discussed above. For example, the Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1) (HHS, 2023a), enhances the ONC Health IT Certification Program by introducing pioneering algorithm transparency requirements for certified health IT and enabling clinical users to assess AI and predictive algorithms for fairness, validity, and safety. However, the HTI-1 rule has come under some scrutiny, in part because there are orders of magnitude more AI tools than there are predictive decision support interventions in regulated EHR systems (Sandalow, 2024); additionally, critical questions remain about the needed standards and specifications for reliable local validation and implementation. And challenges with the EHR certification program, such as self-attestation, periodic recertification, and post-market surveillance, are relevant to AI and will require careful consideration (Ratwani et al., 2024). Meanwhile, multiple parties are endeavoring to address risk by establishing certification standards and assurance procedures for currently unregulated health AI (Shah et al., 2024). Establishing the evidence to support such standards presents one of the greatest challenges to moving forward in certification or assurance programs for health AI.

In the context of planned behavioral theory, there is much to learn from the journey taken to make EHR technology commonplace in the United States. Box 2-1 highlights key lessons. As with the EHR, rigorous evidence is needed to support the use of AI to solve myriad problems to foster multidisciplinary stakeholder understanding of the rationale for the use of AI and to create the

BOX 2-1

Applying Lessons from EHR Adoption to Health AI Implementation

Understand the Rationale for Adopting Health Care AI

- Use evidence to influence the national dialogue
  - Change norms and expectations
- Support behavioral self-control
  - Increase self-efficacy about use of AI
  - Improve understanding about AI among the health care workforce
  - Identify and ameliorate barriers to implementation

Translate Intent into Practice

- Promote standards and certification
- Provide technical support and assistive funding

necessary subjective norms and standards for its adoption and ongoing use. And, similar to EHR implementation, promoting AI adoption will require training and education to support perceived behavioral control, including technical understanding as well as workflow integration. However, while existing regulations and implementation of EHRs may provide a foundation for adoption of AI, some features of AI implementation, such as issues with model sustainability, will require novel approaches to ensure equity, safety, privacy, and usability over time. Aligning the industry to solve these challenges is the goal of the AICC framework.

LEARNING HEALTH SYSTEM (LHS) CONTEXT

In 2006 the IOM (now the NAM) undertook an initiative to identify necessary actions to expand the evidence base for medical decision making, with the dual objectives of ensuring that necessary care is delivered and unnecessary care not delivered (IOM, 2007). A standing group of senior public and private health organization leaders was established with an IOM Charter that developed the concept and definitional parameters of progress toward a continuously LHS: "one in which science, informatics, incentives, and culture are aligned for continuous improvement, innovation, and equity—with best practices and discovery seamlessly embedded in the delivery process, individuals and families as active participants in all elements, and new knowledge generated as an integral by-product of the delivery experience" (NAM, n.d.). See Box 2-2.

However, inadequate progress has been made; central issues identified in the IOM report—evidence not being made available at point of care (Nilsen et al., 2024) and evidence not keeping pace with scientific and technological advances—persist and warrant application of complexity science (Braithwaite et al., 2020).

Recent innovations in AI technologies represent a significant opportunity to ameliorate these challenges. AI can support both basic medical science (such as molecule discovery or drug design) and implementation research with processes including literature reviews, study design, participant recruitment, data analysis, and generative modeling, making it possible to reduce the time from study execution to publication. Further closing the gap from publication to clinical practice, though it currently has limitations, AI shows promise in synthesizing systematic reviews (Nilsen et al., 2024) which may support more rapid updates to evidence-based practice guidelines. AI can also integrate new studies with existing research to produce rapid updates to evidence-based guidelines and provide more personalized treatment recommendations at the point of care. With its capacity to handle extremely large, distributed, and diverse datasets in real time, AI holds promise for improving evidence-based practice (EBP) through improved evidence generation, clinical decision support, and patient shared decision making, three essential components of EBP (Nilsen et al., 2024). AI also holds promise to advance the LHS, because at its core the LHS is predicated on the data generated in patient encounters as well as other sources such as wearables and public health surveillance, creating new information and learning to benefit all. See Figure 2-4.

However, AI introduces new risks and may also amplify existing risks. Considerations including data quality, bias, lack of explainability and transparency in some settings, and performance drift over time; these topics will be detailed further in this work and could yield results that are misaligned with the goals and

BOX 2-2

The Vision of the Learning Health System

Is "one in which science, informatics, incentives, and culture are aligned for continuous improvement, innovation, and equity—with best practices and discovery seamlessly embedded in the delivery process, individuals and families as active participants in all elements, and new knowledge generated as an integral by-product of the delivery experience" (NAM, n.d.).

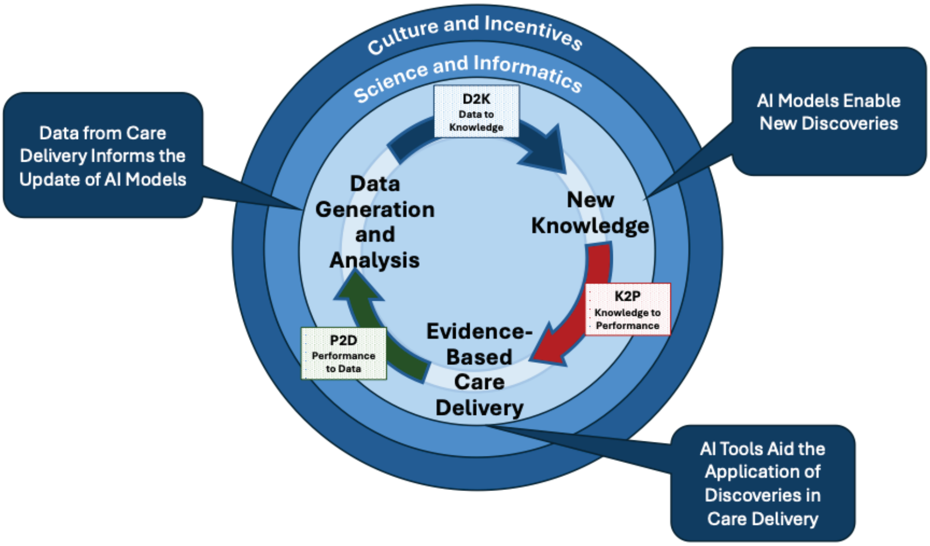

FIGURE 2-4 | Application of AI throughout the Learning Health System cycle.

shared commitments of the LHS (McGinnis et al., 2024). However, many of the risks of AI themselves can be evaluated, characterized, and mitigated over time within the context of the LHS paradigm. In fact, AI systems that can monitor and support other AI systems are seen by many as paramount to avoiding harm. AI may usher in a new era of iterative evidence generation and application, supporting the long-awaited realization of the LHS.

3
AI CODE OF CONDUCT FRAMEWORK

SUMMARY OF PROCESS FOR DRAFTING PRINCIPLES AND COMMITMENTS

The National Academy of Medicine (NAM) AI Code of Conduct (AICC) framework described here was developed to align the field and catalyze action to ensure that the potential of artificial intelligence (AI) in health, health care, and biomedical science is realized. The expectation is not of wholesale adoption of the AICC framework across the industry. Rather, the AICC framework is intended to serve as a touchstone for organizations and groups developing their own considerations and approaches for inclusion and alignment when assessing internal guidance for completeness in their specific context, thereby advancing trust and minimizing the likelihood of actors across the field working at cross-purposes. The drafting of the AICC framework, consisting of the Code Principles and Code Commitments, was intentionally aligned with prior efforts of academics, industry, and governmental agencies to elucidate guiding principles for responsible AI in the health care and biomedical science domains. As detailed in the draft AICC framework (Adams et al., 2024), a landscape review was conducted to identify commonalities published to solicit public and industry comment in existing frameworks as well as critical gaps. Peer-reviewed literature highlighting considerations for responsible AI published between 2018 and 2023 was examined as was guidance on responsible AI produced by medical specialty societies and federal and international policy makers. Additionally, key informant interviews were held with industry experts to solicit input on essential components for the AICC framework.

In the context of the NAM Learning Health System Shared Commitments, the Code Principles reflect the values and norms to be applied in health AI governance to promote trust while ensuring the benefits and mitigating the risks associated with AI in health, health care, and biomedical science. Consistent with the complex adaptive systems theory, which posits that a small set of simple rules can result in system-level change in complex systems (IOM, 2001), the Code Commitments are a set of decision-oriented rubrics intended to support the application of the Code Principles in practice and to guide the behavior of individuals, organizations, and communities, as well as local, national, and transnational agencies operating in complex systems. Table 3-1 presents the relationship between the Code Principles and Code Commitments, reflecting the distillation of the Code Principles into broadly applicable guidance for decision making across impacted parties and across the AI lifecycle. Cells shaded in green reflect an alignment between the Code Principles and the Code Commitments. For example, Commitment 1, Protect and advance human health and human connection as the primary aims, addresses the Code Principle "Secure" in that ensuring privacy and security of health data is integral to protecting human health and human connection, as a breach of data security could result in both breakdowns in mental well-being (health) and in trust (human connection).

SUMMARY OF PUBLIC COMMENT

The AI Code of Conduct draft for public comment, including the Code Principles and Code Commitments was published by NAM in April 2024 (Adams et al., 2024). Public comment was solicited on release, and through an online survey; additionally, presentations on the AICC framework were requested by federal agencies, industry coalitions, collaboratives, and professional association meetings, where verbal feedback was provided to the NAM. Reactions to the AICC framework were provided by a diverse group including federal agencies, researchers, clinicians, AI developers, patients and family representatives, health product manufacturers, ethics and equity experts, standards-setting bodies, and care delivery systems.

Feedback was provided to the NAM staff verbally and in writing through industry events, and, for individuals, an online survey instrument. The survey section provided a 3-point scale (Yes, Somewhat, No) response options to reflect agreement that the Code Principles and Code Commitments engendered various qualities, including clarity, relevance, completeness, conciseness, robustness, and adaptability. Elaboration was also encouraged.

The vast majority of respondents found both the Code Principles and Code Commitments to be relevant, clear, and concise. Likewise, a smaller group, though still a majority, found the Code Principles and Code Commitments to be complete,

TABLE 3-1 | Crosswalk of Draft AICC Code Principles and Commitments

		AICC Code Commitments					
		Protect and advance human health and human connection as the primary aims	Ensure equitable distribution of benefit and risk for all	Engage people as partners with agency in every stage of the lifecycle	Renew the moral well-being and sense of shared purpose to the health care workforce	Monitor and openly and comprehensibly share methods and evidence of AI's performance and impact on health and safety	Innovate, adopt, collaboratively learn, continuously improve, and advance the standard of practice
AICC Code Principles/ LHS Shared Commitments	Engaged						
	Safe						
	Effective						
	Equitable						
	Efficient						
	Accessible						
	Transparent						
	Accountable						
	Secure						
	Adaptive						

NOTE: Cells shaded in green reflect an alignment between the Code Principles and the Code Commitments.

robust and adaptable. Despite high levels of positivity about the conceptual framework, only about half of respondents felt that the set of Code Principles and Code Commitments was sufficient to focus attention and action on issues anticipated to ensure that use of AI in health and health care optimally advances the human condition. This result was expected as outlined in the publication of the draft framework, which was proposed as a "starting point for real-time decision making and detailed implementation plans to promote the responsible use of AI" (Adams et al., 2024, p. 6). Also outlined in that publication, the plan for the second phase of the project, was to take the Code Commitments to the next level of granularity with working groups describing their accountabilities and essential collaborative actions and activities mapped throughout the AI lifecycle and aligned to stakeholder perspectives and responsibilities. The output from these working groups, an online survey, post-presentation feedback, and ongoing expert recommendations, yielded additional edits for clarity and emphasis of the Code Principles and Code Commitments (see Tables 3-2 and 3-3). From the project outstet, the authors underscored that the AICC framework will require ongoing review and, as necessary, revision based on experience with application of the framework as well as changes mandated by advancing AI technologies and the evolution of governance capabilities.

TABLE 3-2 | Updated AICC Code Principles

AICC Principles
Engaged: Understanding, expressing, and prioritizing the needs, preferences, goals of people, and the related implications throughout the AI lifecycle
Safe: Attendance to and continuous vigilance and controls for potentially harmful consequences from the application of AI in health and medicine for individuals and population groups
Effective: Application proven to achieve the intended improvement in personal health and the human condition, in the context of established ethical principles
Equitable: Application accompanied by proof of appropriate steps to ensure fair and unbiased development and access to AI-associated benefits and risk mitigation measures
Efficient: Development and use of AI that results in reductions in resources to achieve improved health outcomes without concomitant adverse impacts on the natural environment
Accessible: Ensuring that seamless stakeholder access and engagement is a core feature of each phase of the AI lifecycle and governance
Transparent: Provision of open, accessible, and understandable information on component AI elements, performance, and their associated outcomes
Accountable: Identifiable and measurable actions taken in the development and use of AI, with clear documentation of benefits and clear controls and accountability for potentially adverse consequences
Secure: Validated procedures to ensure privacy and security, as health data sources are better positioned as a fully protected core utility for the common good, including use of AI for continuous learning and improvement
Adaptive: Assurance that the accountability framework will deliver ongoing information on the results of AI application, for use as required for continuous learning and improvement in health, health care, biomedical science, and, ultimately, the human condition

TABLE 3-3 | Updated AICC Code Commitments

AICC Commitments
Advance Humanity: Protect and advance human health and connection as the primary aims
Ensure Equity: Ensure equitable distribution of benefit and risk for all
Engage Impacted Individuals: Engage people as partners with agency in every stage of the lifecycle
Improve Workforce Well-Being: Renew the moral well-being and sense of shared purpose in the health care workforce
Monitor Performance: Monitor and openly and comprehensibly share methods and evidence of AI's performance and impact on health and safety
Innovate and Learn: Innovate with scalable design, adopt, collaboratively learn, continuously improve, and advance the standard of practice

4
HEALTH AND HEALTH CARE AI LIFECYCLE

As part of the effort to translate the Code Principles and Code Commitments into actionable, real-world practice, it was important to provide a framework for artificial intelligence (AI) that could be referenced throughout this document that encompasses the full lifecycle of activities relevant to the consideration, development, implementation, and sustainment of these technologies. While software development processes have been subject to research and improvement science since the 1980s, there remain multiple development cycle models proposed that provide differing levels of granularity and emphasis relative to their purpose, including those developed for AI. For these reasons, a review of the literature for software development lifecycles with special attention to those proposed for AI was conducted. Well-known lifecycles were described, evaluated, and compared to establish a lifecycle definition for this work that could be used throughout.

"A software process can be defined as the coherent set of policies, organizational structures, technologies, procedures, and artifacts that are needed to conceive, develop, deploy, and maintain a software product" (Fuggetta, 2000). The most basic AI lifecycle model identifies three stages: design, development and deployment (U.S. General Services Administration, n.d.). Others provide more granular detail on certain subprocesses of interest. For example, in the National Academy of Medicine (NAM) seminal publication on AI, *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril* (Matheny et al., 2020), the design phase (which the authors felt had been inadequately addressed in prior literature) was broken down into the following subsections: "(1) identify or re-assess needs, (2) describe existing workflows, (3) define the desired target state, (4) acquire or develop AI system, (5) implement AI system in target setting, (6) monitor ongoing performance, and (7) maintain, update, or de-implement" (Salwei and Carayon, 2022). And the Organisation for Economic Co-operation and Development (OECD) placed significant emphasis on the development stage with substages that included collecting and processing data, building and/or adapting model(s), and testing, evaluation, verification, and validation (OECD, 2024). Other models should be acknowledged (Data Science Process Alliance, n.d.; De Silva and Alahakoon, 2022; Shearer, 2000); however, the author group did not feel that they added key elements beyond the proposed AI Code of Conduct (AICC) model that were needed for this work. The prior NAM and OECD framework (in blue) as well as the AI lifecycle to be used throughout this document were aligned together and shown in Table 4-1.

As part of the review, harmonization, and gap closure process, the author group also reviewed the definitions of each of the stages and substages from a number of prior frameworks to establish definitions for each of the AICC AI lifecycle categories. A summary of these definitions is provided in Table 4-2.

DIFFERENCES BETWEEN AI DEVELOPMENT AND THE STANDARD APPLICATION DEVELOPMENT LIFECYCLE

It is important to note that AI lifecycle considerations have some significant conceptual, technical, and operational differences relative to those for traditional technologies. Here, we seek to generalize considerations broadly for both discriminative and generative AI rather than discuss specific nuances between them. One way to understand these differences is to compare them across stages in the health AI lifecycle.

The *design* of health care AI applications involves a deep understanding of clinical workflows, data, and specific health care needs (Greenhalgh et al., 2017) and is more nuanced than traditional health information applications (e.g., electronic health record [EHR] billing tools), identifying precise problems amenable to AI solutions (e.g., detecting breast cancer). It also involves deeper collaboration across a range of stakeholders including clinicians, patients, family and community members, and implementation and data scientists and practitioners (Hogg et al., 2023; Scott et al., 2021). One example is the definition of use cases and objectives, including specifying how AI will improve patient outcomes, assist with diagnostics, or optimize workflows. While any technology tool has stated goals and objectives for implementation and use, for health AI it is more challenging to ensure that the design goals are met due to training data that may result in biased operation, or produce other unexpected behavior, or not allow transparency or explainability in situations where it is needed (Ferrara, 2023). Early planning for health care AI must also address ethical issues and privacy concerns from the outset. This includes considerations around informed consent, data anonymization, and the potential for algorithmic biases (Khalid et al., 2023).

TABLE 4-1 | A Comparison of Relevant AI Lifecycles with Stage and Content Alignment

Source	AI Lifecycle Stages									
US GSA	Design				Develop			Deploy		
Hope, Hype, Promise, and Peril	Identify or Re-assess Needs	Describe Existing Workflow	Define Desired Target State	System Design and Planning	Acquire or Develop AI System			Implement AI System in Target Setting	Maintain, Update, or De-implement	
OECD	Plan and Design				Collect and Process Data	Build and/ or Adapt Model	Test, Evaluate, Verify, and Validate	Make Available for Use/ Deploy	Operate and Monitor	Retire/ Decommission
AICC	Problem Scoping, System Design, and Planning				Data Acquisition, Management, and Linkage	Model Development, Testing, and Procurement		Implement and Scale	Post Implementation Monitoring, Feedback Systems, and Decommissioning	

NOTES: The lifecycle used for the AICC project and in this document is in yellow at the bottom. AICC = AI Code of Conduct; OECD = Organisation for Economic Co-operation and Development; US GSA = U.S. General Services Administration.

TABLE 4-2 | Description of the AI Development Lifecycle

AICC AI Lifecycle Stage	Description
Design: problem scoping, system design, and planning	In this phase, the goals and stakeholders of the AI system are identified as are the data, tools, and technologies needed to achieve those goals. Standard project management approaches including timelines, milestones, and proactive risk identification and mitigation strategies are employed. This includes use cases and incorporation of user-centered design principles regarding engagement of end users and defining standardized, measurable metrics for successful implementation and outcome evaluation. These metrics refer not only to statistical measures of model success, but also of clinical and economic outcome metrics. Planning for seamless data collection for metrics is also included in this phase as is planning for interoperability and system scalability.
Data Management: data acquisition, management, and linkage	In this phase, data needs identified during planning are addressed. Data may be directly extracted from existing systems or developed using synthetic techniques (with due consideration of the associated benefits and limitations). These data must be securely transmitted and stored with appropriate access controls. The data are assessed for integrity, and transformed to meet model requirements, including assurance for data quality characterization as well as representativeness and generalizability of the data to ensure that they are fit for intended purposes and application to diverse populations. As applicable, datasets may be linked to create a comprehensive repository for the AI system.
Development: model development, testing, and procurement	This phase includes the process of developing new AI models and/or procuring existing AI systems and then adapting them to the local environment. This phase is iterative until the AI system performs to the goals and metrics set out in the planning phase, which includes ongoing algorithm development/training, testing, and refining until the system demonstrates the required accuracy, robustness, and sub-population equity.
Implementation: implementation and scaling	In this phase, AI systems are incorporated into the desired workflow through a user-centered design process to ensure that the predefined goals are met and performance in the setting of use are achieved. Typically, this involves initial deployment in a test environment using real-world data, with monitoring and acceptance testing occurring before implementation in live systems, whereupon additional assessment of model performance is completed.
Maintenance: post-implementation monitoring and feedback systems and decommissioning	The setting in which AI systems are deployed may change in a number of ways. The clinical care processes and data collected may change over time, and the AI tool itself may change the workflow and outcomes it is embedded within. For these reasons, AI systems may produce variable outputs, which may be desirable or may reduce accuracy over time and even result in harm. As such, monitoring of AI system results, including the use of feedback systems, is essential to ensure that the systems are achieving their stated goals and not causing harm. Decommissioning is also critical to consider, which includes removing systems that are not performing as expected or are no longer required or desired.

Digital health applications might not face these ethical challenges to the same extent. The design phase must consider regulatory requirements specific to health care AI. This involves understanding evolving guidelines and requirements from federal agencies and governing bodies and incorporating these into the design process. It is also important to consider cognitive heuristic biases that may be generated from how AI is integrated into workflow, and mechanisms to expose and prevent these biases are important (Jabbour et al., 2023). Unlike rule-based systems, AI solutions may be designed to augment users, requiring greater clarity about roles and accountabilities as well as the associated mechanisms to track and report on feedback loops. Finally, interoperability and scalability require careful consideration and planning during the design phase of AI systems (Oikonomou and Khera, 2024). Of the utmost importance during the design phase, developers, in collaboration with end users and recipients of AI (e.g., patients), must plan for comprehensive evaluation of AI tools and products throughout the entirety of the AI lifecycle. As outlined in the proposed IMPACTS Framework, the evaluation should include the assessment of Integration; Monitoring, governance and accountability; Performance quality metrics; Acceptability trust and training; Cost and economic evaluation; Technology safety and transparency; and Scalability and impact (Jacob et al., 2025). This model addresses the widely used approach of assessing the statistical accuracy of a model, but also includes but is not limited to issues such as interoperability and workflow integration, data privacy and security, usability, safety, clinical effectiveness, clinical efficiency, and clinical utility.

Data management in AI systems differs from digital health application development in that doing so requires large, high-quality datasets for training and validation. Ensuring data quality and mitigating bias are critical during model development (Ahmed et al., 2023a). To avoid creating new or exacerbating existing biases, AI models require datasets that are diverse and representative.

AI systems *development* also differs from digital health system development in the use of advanced machine learning and related AI techniques. Once developed, health care AI models require rigorous validation and testing to ensure their accuracy, reliability, and generalizability, and applicability across different local patient populations. This involves comparing AI outputs with expert clinical judgments and assessing performance metrics such as sensitivity, specificity, and predictive value. AI models often require iterative development based on feedback from clinical trials or pilot studies (Siontis et al., 2021). This iterative process helps to refine the model and improve its performance before full-scale deployment. Developing AI models also involves promoting and facilitating transparency and explainability in the design, inputs, processes, and outputs where possible, sometimes through tools and techniques used in parallel to the core AI. This is especially important in health care contexts where the ability to comprehend the rationale for AI-generated guidance is crucial for clinical decision making (Albahri et al., 2023). Traditional digital health systems might not face such stringent requirements regarding data bias, have less intensive validation processes, have a more linear development process, and typically are much easier to explain, which has led to variability in AI development evaluations (Tornero-Costa et al., 2023).

Additionally, procurement of health AI systems often requires specialized expertise to evaluate appropriateness to a given setting. Unlike digital health solutions, assessing an AI application involves understanding its underlying algorithmic performance, data requirements, and potential biases, all within the context of use. As such, procurement teams may need to include data scientists or AI specialists in addition to clinical, administrative, or technology expertise for traditional health information procurement. Regulatory compliance is another factor in this phase, as health AI applications and their use are subject to evolving regulatory requirements. Procurement processes must also ensure that data privacy and security measures are robust from technological, legal, and contractual perspectives. While these issues are present with digital health procurement, the changing regulatory landscape and the expansion in the quantity of data used in AI systems require additional attention in AI procurements.

Implementation also presents important differences for health AI solutions which often need to be integrated with existing EHRs or other digital health information systems, and often in ways that are more complex and can change over time due to the evolving nature of AI algorithmic solutions (Greenhalgh et al., 2017). AI applications often require customization and calibration based on the specific needs of the health care setting (Brady et al., 2024). This can involve fine-tuning algorithms to local data and practices, which is less common with digital health applications. During implementation of health AI, it is important to assess workflow changes, decision support, and the performance of the AI-enabled clinical team. Implementing AI tools may also require additional training for health care professionals to understand and effectively use the AI system.

Once implemented, AI applications require ongoing substantial *maintenance*. Indeed, continuous monitoring of AI applications, or algorithmovigilance, is crucial to ensure their accuracy and effectiveness (Embí, 2021), including assessment of real-world outcome performance, as planned in the design phase. Unlike previous IT applications, AI systems and their effects and impacts might evolve over time as they are exposed to new data and environments, so ongoing performance evaluation is necessary to detect and correct issues (Davis et al., 2017). AI applications need to be regularly assessed for biases and accuracy (Davis et al., 2024). This is especially important in health care where inequities in data can lead to or exacerbate existing disparities in health outcomes. Health AI systems may require regular updates and retraining as they change or "drift," to adapt to new data, variations in patient populations, and changes in medical knowledge. Finally, monitoring health AI includes ensuring ethical use and addressing concerns about transparency, accountability, and consent in ways that might differ from earlier digital health applications.

In summary, health AI applications share many similarities to digital health applications, but they also differ in important ways. Key differences are depicted in Figure 4-1. In the design phase there must be a focus on understanding clinical needs, defining specific use cases, addressing ethical and privacy issues, navigating regulatory requirements, and planning for data management. Development of health AI involves selecting and developing algorithms, ensuring data quality and bias mitigation, rigorous model validation, iterative development, and striving to promote explainability and transparency. And, procurement requires specialized expertise, regulatory compliance, and careful consideration of data handling. Implementation involves complex integration, extensive training, and customization. Finally, maintenance focuses on performance accuracy, bias, fairness, adaptability, and ethical concerns. These differences reflect the challenges and considerations associated with health AI applications, necessitating tailored approaches compared to earlier digital health solutions.

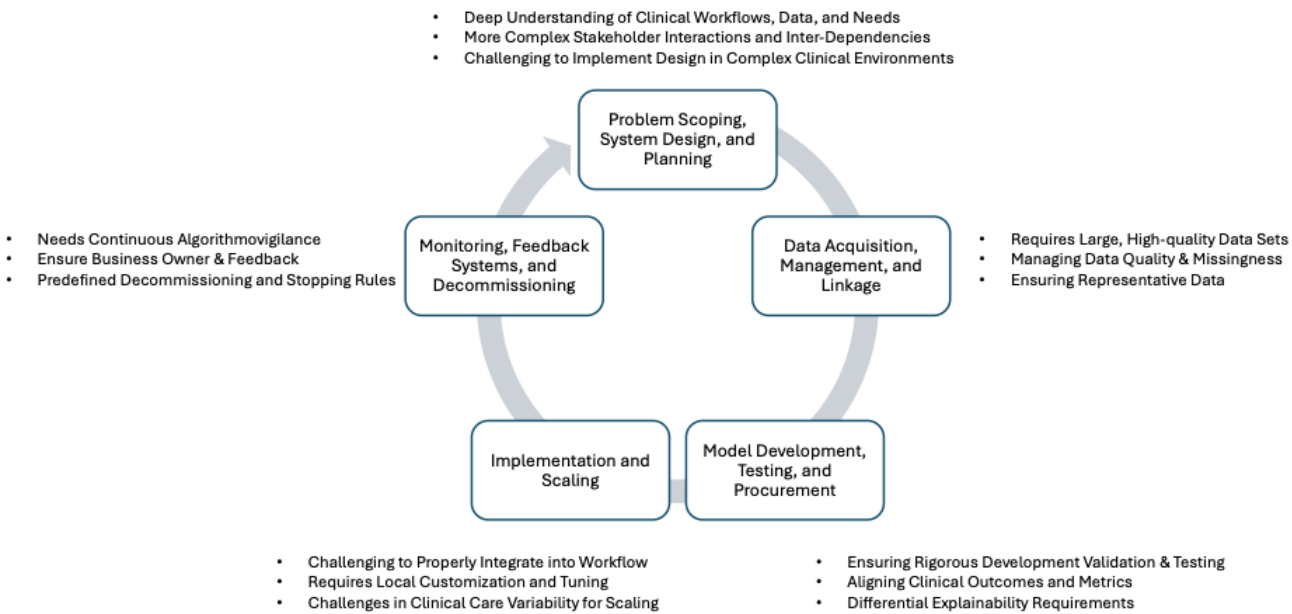

FIGURE 4-1 | Some key differences between AI lifecycle development and standard application development.

5
TRANSLATION OF THE AI CODE OF CONDUCT FRAMEWORK TO REAL-WORLD APPLICATIONS

A central goal of establishing the AI Code of Conduct (AICC) framework is to ensure that the benefits of artificial intelligence (AI) in health, health care, and biomedical science are realized, and the risks mitigated. To reflect an intention of field alignment rather than a one-size-fits-all set of rules, the AICC Code Principles and Code Commitments are intentionally high level, providing guideposts for actors in the health, health care, and biomedical science system. However, to realize the goal of improved health of the U.S. population using AI, it is essential to translate the AICC Commitments into actionable, real-world practice, bringing the AICC Commitments to the next level of granularity by providing context and examples which are intended to be used to support the development of more comprehensive implementation guides by industry sectors and individual organizations. To that end, the AICC steering committee members convened expert working groups to consider and characterize real-world activities and stakeholder perspectives, responsibilities, and needs in the context of the Code Commitments and the AICC AI lifecycle. Table 5-1 provides description of the stakeholder groups directly involved in or subject to the development, use, and evaluation of health AI. *It should be noted that some of the content is duplicated across stakeholder perspectives as it is anticipated that readers may choose to focus on stakeholder perspective content most relevant to themselves.*

AI DEVELOPERS' PERSPECTIVE

Developers shoulder immense responsibility when creating AI solutions that impact human health, and conceptually these communities exist as both individuals and organizations. The AICC Commitments warrant consideration at every stage of the AI lifecycle, including algorithm development, and deployment

TABLE 5-1 | Description of Key Stakeholder Groups

Stakeholder Group	Brief Description/Definition
AI Developers	This perspective includes individuals who write and engineer AI models for health care applications, including those who use no-code or low-code tools or approaches to develop AI models for health care applications. This also includes companies and organizations that develop AI as stand-alone software or embedded within medical devices or larger health care solutions. Lastly, this includes organizations that are not only developing algorithms directly, but also those developing technologies directly integrated with AI, such as physical medical devices, biotechnology companies, and electronic health record vendors. These organizations span the continuum from small start-up companies and initiatives to well-established large companies that cover a full spectrum of innovation, development, and iterative improvement of AI tools and capabilities for use in health care for or by patients.
Researchers	This perspective includes individuals and organizations that seek to innovate and develop new capacities. The research sector, by its nature, seeks to innovate and develop new capacities not only throughout each stage of the AI lifecycle, but also within relevant foundational theories, concepts, and systems in data science, technology, ethics, anthropology, and many more. As individuals, researchers include people who seek to apply the scientific method to rigorously conceptualize, evaluate, and replicate novel theories, frameworks, systems, and applications. Research organizations include federal agencies responsible for research (National Science Foundation [NSF], the National Institutes of Health [NIH], and the Agency for Healthcare Research and Quality [AHRQ]) and businesses that seek to gather individual researchers to drive innovation and discovery as one of their primary missions, such as academic institutions and research divisions of commercial companies, including pharmaceutical manufacturers.
Health Systems and Payors	A health system is a collection of people and entities that delivers health care services to meet the health needs of populations. This includes all the resources, services, and capacities that are involved in the conduct of health care delivery. A health care payor is any entity that pays for services rendered by a health care provider, and by extension, health systems. This could be a private employer, a commercial insurance company, or a government program.
Patients	Patients are the consumers of health care and the eventual recipients, directly or indirectly, of public health and health care delivery. They are critical stakeholders in all aspects of health AI, as the end goal of AI in this domain is to support human health and well-being. Patient advocacy organizations are also a critical stakeholder in this group, as they serve to support patients and caregivers to champion their health needs, to educate them on health AI's opportunities, challenges, and concerns, and to provide direct representation of patient groups to inform policy and improve research design, ethics, and logistical barriers, among others (Patterson et al., 2023).

TABLE 5-1 | Continued

Stakeholder Group	Brief Description/Definition
Federal Agencies	Federal health agencies and offices (collectively, "agencies") have various authorities, tools, programs, and incentives to encourage and even require specific actions to ensure responsible and ethical adoption of AI in health. Many, but not all, of these agencies are within the Department of Health and Human Services. Some notable agencies with regulatory functions include the U.S. Food and Drug Administration, the Centers for Medicare & Medicaid Services, the Centers for Disease Control and Prevention, and the Assistant Secretary for Technology Policy and Office of the National Coordinator for Health Information Technology. This also includes agencies conducting health care delivery, such as the Veterans Health Administration, and the Indian Health Service, as well as those primarily responsible for research support, such as NSF, NIH, and AHRQ.

process, prioritizing patient alignment, security, and ethical considerations over commercial interests or individual gains. Key considerations for action mapped to the Commitments, as applicable, are discussed below. In particular, it is essential that organizations recognize potential conflicts between developing profitable AI and aligning with the Commitments and seek to prioritize the Commitments while still meeting business needs.

To that end, as described in the AI Development Lifecycle section of this chapter, developers should engage in a comprehensive impact assessment (Jacob et al., 2025) for new and existing products. Additionally, near-, short-, and long-term potential benefits and harms across the various involved parties in the value chain throughout the AI lifecycle should be identified, weighed, and mitigated where possible (Attard-Frost and Widder, 2024), and transparently shared with users of AI.

Commitment 1: Advance Humanity

Developers translate the societal, cultural, and individual goals of health and health care from other AI stakeholder groups into technical implementations that prioritize human agency. Developers also can play a strong role in fostering human connections by carefully designing human-computer interactions to optimize human communication. In health and health care, technical and algorithmic design choices will best advance humanity when aligned directly with individual goals and family priorities and when developers proactively work to mitigate misalignment in that respect (The Light Collective, 2024). In addition, it is important that developers consider how AI may be used downstream to design mechanisms for patients and users to be aware of what (and how) the AI is operating, making transparent whether it is appropriate to be used with regard to each observation, patient, or use.

Commitment 2: Ensure Equity

Developers play a central role in ensuring that bias is assessed and mitigated, and fairness is promoted within the context of use of AI applications and are also responsible for translating the societal and cultural concepts of equity from users and receipts of AI use into technical implementations. There are ethical considerations for developers throughout the AI implementation lifecycle. As developers may not have formal ethical training and in most cases will not be able to represent the interests, culture, and preferences of end users and patients, inclusion of these groups as well as experts in ethics and equity can help guide the development process, ensuring that an AI solutions' primary aim is to advance human health. This is because it is critical to proactively identify starting with the initial design and conception, and this can be done by incorporating diverse perspectives, ensuring representative datasets, and implementing fairness metrics that are evaluated up front as well as on an ongoing basis (Consumer Technology Association, 2023; Echo Wang et al., 2022).

One of the cornerstones of ensuring equity and fairness and mitigating bias is for developers to prioritize and use data in AI development that are provided within clear governance frameworks that respect patient privacy, promote data equity, and seek to ensure high-quality, diverse, equitable, and representative datasets. It is also important for developers to consider and address AI models' technical proficiency and security, as well as ethical soundness and clinical relevance—partnering with stakeholders managing the environment of use, such as health care systems, and leveraging insights from federal agencies. Last, developers can require, facilitate, and recommend to users a thorough risk assessment and benefit measurement approach that includes continuous monitoring and evaluation of AI systems to identify and mitigate potential risks while optimizing benefits for all parties that include health care providers, patients, and others.

Commitment 3: Engage Impacted Individuals

Moving beyond technology, developers must embrace a people-centered approach that prioritizes collaboration, transparency, and shared decision making. Engaging people as partners with agency throughout the AI lifecycle is crucial for developing responsible and effective AI solutions. For developers, this directly and clearly translates to identifying and promoting engagement, representation, input, and feedback from any constituent group involved in the conceptualization, development, use, or receipt of actions or decisions from AI use throughout the entire AI lifecycle. Collaborating with these groups (e.g., patients, clinicians, and administrators) and seeking their input and feedback is foundational to ensure solutions align with their needs and expectations. These needs include concepts of fiduciary responsibility, legal liability, goals of care or desired outcomes, and concerns or risk tolerance. It should also be noted that for any AI systems where individuals are patients and recipients of AI output, prioritizing their representation throughout the development process takes on additional urgency, as does actively seeking their input and ensuring that their interests are advanced and protected.

It is also important for developers to consider the conflicting needs of multiple stakeholders, as AI products will provide the greatest benefit to humanity only if they balance the needs and preferences of multiple parties. This is of particular importance when users and the recipients of AI outputs or recommendations are not the same individual. While openly acknowledging and addressing potential conflicts among all stakeholders, the prime driver must remain the prioritization of equity, well-being, and health care advancement. This includes embracing the tension between different needs as a catalyst for critical reflection, innovation, and improvement. Embracing a continuous improvement mindset to ensure solutions remain effective and responsive is needed.

Lastly, developers should consider that they are themselves important stakeholders in the overall AI ecosystem, and that they should seek to contribute and engage as stakeholders in upstream activities that impact the AI implementation lifecycle broadly. Developers have vast experience to share in developing frameworks, standards, foundational sociotechnical constructs, and their participation in these activities would be valuable for building baseline methods and practices for the whole ecosystem.

Commitment 4: Improve Workforce Well-being

Developers play an important role in conceptualizing, developing, and implementing software solutions that improve workflows and reduce inefficiency and cognitive burden. This group plays a critical role of translation in understanding the objectives and needs of health care users and recipients of AI tools to optimize the richness, quality, and safety of human interactions in health care delivery. To accomplish this, developers must work with and engage the workforce of both health professionals themselves and health care well-being experts who can translate knowledge and best practices through user preferences. This can result in valuable input about the impact that AI solutions might have on clinical workflows and practitioner morale and provide opportunities for optimization. Workflow may be optimized by aligning clinician preferences regarding when, where, and how outputs of the AI solutions should be displayed and shared with the internal technology teams for optimal integration into workflows (Chen et al., 2022).

Lastly, a critical challenge in ensuring satisfaction for any workflow process change is to appropriately support communication and education in its use and provide proactive plans for implementation by clinical and operational staff. While this is not generally performed by the developer, including guidelines and best practices for how organizations may most effectively implement their AI solution is likely to improve adoption and user satisfaction. These guidelines should also include any required clinician and patient disclosure statements about the use of the AI solutions, and those must be understandable, actionable, and clearly defined.

Commitment 5: Monitor Performance

Transparency builds trust and ensures responsible AI development. A foundational issue for developers is the need to intentionally design in a way that will increase the user's trust in AI. There are several ways in which developers can promote transparency and support the maintenance of performance over time. This includes committing to monitoring AI performance and sharing information about its impact on health and safety.

First, developers should maintain appropriate transparency regarding data sources, including demographics, location, and time period of data collection so that trust can be built on the quality of the solution and the applicability to the current context (system and patient population). Understanding the data sources for AI are critical in understanding the allowable context of use and establishing appropriate performance metrics.

Developers can play an important role in promoting trust and facilitating informed decision making by relevant stakeholders by proactively developing monitoring platforms and openly disclosing product performance metrics and any specific implementation requirements, as well as feedback loops and mechanisms to support post-implementation monitoring, recalibration, and decommissioning. Developers should promote processes for monitoring, feedback, and improvement, drawing from various disciplines and expertise to ensure diverse perspectives are considered (Davis et al., 2024). Robust post-market monitoring systems with clear channels for user feedback (including patients and providers) and reporting of concerns (Saria, 2022; Vasey et al., 2022) provide a mechanism to ensure that health AI systems continue to achieve their stated goals over time.

During the development process and initial use piloting phases, root cause analyses of errors associated with a harm should be conducted without fear of retaliation because the process is critical to allow developers to learn and improve. Existing examples include patient safety organizations and coordinated vulnerability disclosure (cybersecurity) (Householder et al., 2017). From these activities, developers should also provide guidelines for characterization of harms and errors as well as remediation process recommendations for anticipated failure modes to downstream users. Additionally, proactive planning can be performed for unintentional information disclosure, considering the needs of affected parties and generating recommendations for tailoring communication strategies accordingly. Lastly, this stakeholder group should collaborate with researchers and federal agencies to promote standards for necessary oversight and support for adapting AI solutions to emerging evidence and shifting patient needs.

Commitment 6: Innovate and Learn

The developer community plays a crucial role in ensuring responsible and ethical AI innovation that enables and enhances clinicians' ability to advance clinical practice and ensuring that AI is intentionally designed for scalability. Ongoing exploration of new approaches and incorporation of diverse perspectives are needed to ensure the robustness and reliability of AI models and solutions (Saria, 2022). This requires active participation, collaboration, and a dedication to continuous learning. Staying at the forefront of AI advancements through exploration, research, and knowledge sharing within their community, developers can contribute to the ongoing improvement of health care delivery. It is critical for developers to actively seek feedback and incorporate insights into new product design as well as iterations of deployed AI models (Saria, 2022; Vasey et al., 2022). Open communication and collaboration with stakeholders will be crucial to navigate potential misalignments and ensure that solutions are ethically sound and beneficial to all.

For most use cases, AI solutions are likely to augment users. In these scenarios, a clear teaming model should consider user roles, collaborations (how user(s) collaborate with and leverage the software to complete a task), and accountabilities (for what each party is accountable). Additionally, interfaces that are transparent and intelligible and enable each party to responsibly perform their intended role are advantageous (Henry et al., 2022).

As larger, more sensitive datasets are used in health AI, it will be paramount to continue to develop methods to support the use of unbiased, real-world data, while ensuring data provenance and carefully considering data linkage to protect patient privacy and align with user groups' interests. Additionally, ongoing investment in the innovation in synthetic data creation is likely to preserve privacy while promoting fair and representative data availability (Li et al., 2023).

Developers should learn from successes and continued challenges as processes for monitoring, feedback, and improvement are implemented, and they should continue to innovate in collaboration with various disciplines and expertise to promote more equitable and fairer algorithmic performance for both statistical and clinical metrics of importance (Davis et al., 2024).

By embracing the AICC Commitments and prioritizing ethical considerations, the developer community can drive responsible AI innovation that empowers health care professionals to advance the standard of clinical practice and improve patient outcomes, while simultaneously building trust among clinicians and the public and advancing their commercial interests. This requires a dedication to understanding user needs, fostering collaboration, ensuring data integrity, and continuously striving for improvement. Through these efforts, developers can contribute to a future where AI plays a transformative role in health care while upholding the highest ethical standards.

RESEARCHERS' PERSPECTIVE

Researchers have a responsibility in both the ethical conceptualization and conduct of research but also in the application of their research. Research has driven many innovations but also can lead, and has led, to significant harms in the absence of adherence to an ethical code, such as that which occurred within the Tuskegee Study of Untreated Syphilis (CDC, n.d.). All research is now governed by "the principles of respect for persons, beneficence, and justice" outlined in the Belmont Report (HHS, 1979). These principles for human subject protection continue to improve and evolve (O'Sullivan et al., 2020). A community-accepted AI code of conduct framework is important in the research domain when considering questions relevant to the design, use, and impact of AI.

In the context of health AI, there are two major areas of relevant research: (1) AI methods and their validation (basic research), and (2) assessment of AI's impact in the lab or in real-world settings (applied research). First, AI tools are increasingly being used to conduct research, supporting researchers in reviewing the literature, drafting portions of manuscripts, and writing analytical code (Koller et al., 2024). Second, researchers are studying how to develop, adapt, implement, and sustain AI systems, tools, and applications to support the goals and objectives of patients, clinicians, and health care organizations. Researchers must continue to adhere to ethical standards, ensure internal and external validity, be transparent regarding assumptions and limitations and assess and mitigate inequitable performance (NIH, 2024), as well as follow open science guidelines. AI presents novel opportunities and challenges, requiring consideration beyond existing ethical frameworks. These issues are presented below, organized within the AICC Commitment framework.

Commitment 1: Advance Humanity

AI findings and tools developed by researchers are increasingly being employed in clinical and non-clinical environments, thereby directly impacting human health and placing a greater need to consider their unintended consequences. Accordingly, potential impacts—good and bad—are of greater magnitude than in research without AI due to its immense scope and breakneck pace of implementation. The concepts of beneficence and need to avoid harm, taken from ethical frameworks, require more robust engagement with patients to develop appropriate methods of shared governance and consent.

Commitment 2: Ensure Equity

There are myriad aspects to equity warranting further research and innovation. Most importantly, researchers should consider equity and inclusion from conceptualization to execution, evaluation of mechanisms for equitable implementation, and considerations for how and when technologies or tools should be retired. Through this lens, key aspects of ensuring equity through research emerge.

Diverse data sources are essential to ensure equitable health AI benefits. Programs, such as Bridge2AI,[1] that enable ethically sourced data to train AI models—before they are applied in care—are crucial to create a virtuous learning cycle. This also includes innovation and appropriate protections in data access and sharing. While requirements are in place for sharing of data used for research, multiple studies have found that a minority of research benefits from shared data. This extends to data for development and use in health AI; additional research is warranted to maximize data synthesis (Gonzales et al., 2023; Guillaudeux et al., 2023), data de-identification and sharing along the themes of societal benefits, distribution of risks, benefits, and burdens, respect for persons, and public trust and engagement (Kalkman et al., 2019).

There are significant challenges in ensuring that AI technologies, once implemented, achieve sustained performance and avoidance of sub-population

[1] See https://bridge2ai.org (accessed June 21, 2024).

bias. This is an important area of innovation need and opportunity to provide methods and workflows to support continuous monitoring for both AI algorithmic performance and data fitness for purpose. All AI should be monitored for performance (including AI-driven monitors), innovation in automation is critical to provide coverage and scalability as health AI becomes more ubiquitous. It is important to understand that monitoring is only one part of the overall approach to equity, as considerations should be made to address bias and equity through appropriate choices in data sources and AI algorithmic design as well.

Lastly, the use of AI is iterative and cyclical in nature, mirroring the Learning Health System (LHS) by ingesting data created by health care processes to generate information and evidence that is fed back into the health care system. While potentially speeding up translation of data to knowledge to practice, this cyclical nature also means that without appropriate intervention, any biases and inequities present in the health care system and resulting data may be perpetuated and potentially magnified by AI applications (Leslie et al., 2021; Suresh and Guttag, 2021). It is important to conceptualize both AI research and downstream use as a virtuous cycle with each activity informing improvements and new directions in the other.

Commitment 3: Engage Impacted Individuals

Efforts in generating evidence regarding new frameworks, methods, and tools to support AI implementation in health care critically need representation from developers, subject-matter experts and end users, ethicists, patients, caregivers, organizational and cultural groups, and legal and policy makers. Intended as examples from the large numbers of stakeholders already noted, provided below are some specific engagement recommendations.

Researchers can *engage federal agencies and non-profit foundations* to advocate for increasing investment in foundational and applied health AI research that is in alignment with emerging best practices and for continued expansion of knowledge and understanding of all aspects of health AI from design through implementation and maintenance.

Researchers can actively *engage with study participants and advocacy groups and communities that represent their interests*—especially those who are under-represented in research—in every stage of AI research, from conceptualization to design, execution, and interpretation and findings. This can include directly embedding patient and community representatives in the conduct of research but also promoting and advocating for patients and communities to be involved in research agenda setting and priorities. Additional research should be done in the science of patient and community engagement to facilitate deeper integration throughout the AI lifecycle and in ways that ensure patient and community perspectives are integral to the process.

Researchers can *engage health systems and payors* to understand the business and clinical needs facing these organizations, and partner with them in the conduct of research to characterize, evaluate, and address challenges and gaps.

Commitment 4: Improve Workforce Well-Being

Research funding is needed to pursue activities that apply AI to reduce administrative burden, to improve human–human connection and communication, and to reduce clinical cognitive burden as well as to improve data quality, completeness, and timeliness. Research is needed to expand the understanding of the sociotechnical aspects of health AI as well as the best practices and standards in workflow integration. Numerous sociotechnical frameworks exist, but efforts to harmonize these learnings and promote maturity in this domain are important (Reddy et al., 2021; Salwei and Carayon, 2022). Researchers have an opportunity to play a leadership role in the development and validation of AI technologies, such as ambient scribes, that can reduce administrative burden, improve workforce–patient communication time, and reduce cognitive workforce burden.

Commitment 5: Monitor Performance

Researchers are poised to make major contributions, in partnership with other stakeholders noted, toward the development and promotion of methods and tools to facilitate the monitoring and sustainment of AI technologies' performance in the context of use. AI algorithms should strive for explainability and plausibility at the model level, with clear assessments of data inputs and performance outputs being regularly conducted for efficacy, bias, equity, and safety to ensure trustworthiness of AI use (Adam et al., 2020; Feng et al., 2022). Researchers can also play a central role in expanding the understanding and capacity to anticipate unintended consequences, to establish best practices to prevent their occurrence, and to develop monitoring capacities to identify and mitigate them when they do occur (Suresh and Guttag, 2021).

Commitment 6: Innovate and Learn

Health care is a large-scale, dynamic system—and the "butterfly" effect of AI interventions can be significant and difficult to anticipate. AI's speed and scope are accelerating, sometimes at odds with the methodological, bias-controlled approach to research (Lorenz, 1972). AI systems are dynamic, more akin to behaviors than fixed output. Enhanced methods to dynamically monitor AI output and up-front checks on model behavior are required to ensure researchers' ability to fulfill the principles of the ethical frameworks, especially in protecting the rights of and avoiding harm to people.

Sharing research about AI can be challenging, but researchers could develop, promote, and adopt reporting frameworks that facilitate transparency and reusability of AI research. Adaptations of frameworks such as FAIR and CARE (Carroll et al., 2021) and protocols for research such as Consolidated Standards of Reporting Trials for AI and Protocol for Development of a Reporting Guideline for AI are warranted to ensure optimal openness of research while considering the impacts on communities, especially those who have faced historical harms from research. These changes enhance the focus on fairness, separation of training and test sets, and suggest new ways to engage affected communities so they can exert more control and shape ethical approaches. Ultimately, algorithmic transparency and the research used to demonstrate its effectiveness will be crucial to achieve trust and enable AI sustainability. And given that private companies may be reticent to share their findings without a strong business case to do so, public funding of AI research will remain essential to developing trust.

In conclusion, the rapidly changing, large-scale nature of AI coupled with the complexity of health care can lead to unforeseen consequences and new requirements of researchers. Researchers have new responsibilities in ensuring the validity and equity of AI applications and implementation, monitoring and oversight of deployed AI systems in health care settings where they may exhibit emergent or unintended behaviors. Equity may be especially affected, as public perception of algorithms' fairness may clash with the reality of algorithmic bias. Ultimately, humans, including researchers, are responsible and accountable for any biases and errors made by AI systems.

HEALTH SYSTEMS' AND PAYORS' PERSPECTIVE

As noted in the introduction of this chapter, the health system is experiencing significant challenges for which they are actively seeking AI solutions. Simultaneously there is a dearth of experience with these new tools and organizational-level governance of them. Yet, these organizations have a leadership role in specifying and promoting the business and clinical needs and requirements for the use of health AI and bear a strong responsibility for ensuring that AI used in care delivery benefits patients equitably and is used in a way that builds trust in the health care system. This includes carefully considering issues including patient privacy, consent, agency, and accountability, addressing legal and financial responsibilities, and overcoming challenges related to the workforce. By implementing strategies designed to support these priorities, health systems and payors demonstrate accountability in their pledge to harness AI's potential effectively, safeguard human connections, mitigate risks, and foster a more inclusive and trustworthy health care environment (Dorr et al., 2023). Outlined below are health systems' and payors' actions that are aligned with and advance the AICC principles and commitments, as well as the AI lifecycle where appropriate.

Commitment 1: Advance Humanity

While health systems and payors typically have patient health as a primary aim in their mission statement, in the context of health AI, they could go further to establish clear aims prioritizing patient health as paramount, and engaging patients and health care professionals as partners to identify key opportunities where AI can enhance human connection within health care settings. Health systems and payors can require that AI technologies that are to be used be adapted to the "local" context of use, be patient-centric, and incorporate appropriate local clinical and technical expertise and patient insights to ensure relevance and effectiveness (Sauerbrei et al., 2023).

Commitment 2: Ensure Equity

It is important that health systems and payors require transparent conduct and reporting of equity impact assessments of health AI to identify and mitigate potential biases or disparities in access and outcomes (Kim et al., 2024). These activities should involve patients and communities in the designing, developing, evaluating, and governing of AI applications, respecting their preferences and values, to ensure that AI technologies are patient-centered and benefit all parties equitably.

One of the largest challenges to ensuring equity in the use of health AI is the development of frameworks, processes, tools, and workflows that are scalable, inexpensive to implement and maintain, and are widely available to health consumers, particularly those in under-resourced environments or settings. The absence of these characteristics will promote inequitable access to health AI and restrict access of these technologies to only large, well-resourced systems. Health systems and payors can promote equity in this axis by selecting AI technologies that express scalable characteristics that make use and implementation easier.

It is important to recognize the historical and systemic sources of mistrust, harm, and bias that health system and payor behaviors have engendered (Brown et al., 2024; CDC, 2022). This stakeholder group should facilitate processes to ensure that AI does not exacerbate or perpetuate these issues by actively working to eliminate actions that erode trust or result in biases or harms within AI systems. One aspect of this could be requirements that health AI selected for use include patient-understandable, culturally appropriate information so that patients and their caregivers can assess AI's potential risks and benefits in context with their care goals. This includes transparency regarding both the AI technologies and the patients' health data used by and generated from such tools. Enhancing privacy (Murdoch, 2021) and consent (Perni et al., 2023) practices is also central to upholding equity and trust in AI applications and requires a multifaceted approach to promote transparency, adaptability, patient involvement, and regulation adherence. This includes establishing robust mechanisms to maintain accountability for, disclose occurrences of, and mitigate AI-based harms to uphold ethical standards and maintain patient trust, and regular review and updating of these policies in response to evolving technology and assessment of performance in the local setting.

Health systems and payors can provide a convening and facilitating role in promoting transparency and balance in the goals and values between health care delivery and patient goals and needs. It is important to establish a multifaceted approach to balancing goals and values among health and health care stakeholders that acknowledges differing values and goals and emphasizes shared benefits. Increasing transparency and accountability for health systems and payors, and information accessibility for patients, is paramount to support these processes.

It is important that health care organizations develop a culture of equity that includes making health equity a strategic priority, creating structures and processes to support equity work, and eliminating structural racism. These initiatives can help ensure equal distribution of benefit and risk for all (Wyatt et al., 2016). This includes aligning the incentives and payment models of patients and health systems with the objectives of improving health outcomes, quality, and equity for patients with a goal of mitigating the potential for AI misuse or financial exploitation.

Data governance, model development, and model implementation practices should promote equity by collecting diverse and representative datasets to ensure the AI system works effectively for all patient groups that the model seeks to benefit (Juhn et al., 2022). This includes regular performance reviews, which are important to identify and address emerging disparities, ensuring that benefits and risks are equitably distributed. In addition, organizations should consider establishing robust mechanisms to maintain accountability for, disclose occurrences of, and mitigate AI-based harms to uphold ethical standards and maintain patient trust.

Lastly, as part of the overall efforts to maintain patient trust and deliver ethical care, it is important to adhere to regulations, standards, and certification for data quality, security, and governance, such as the Health Insurance Portability and Accountability Act (HIPAA) (HHS, 1996); the International Organization for Standards (ISO) (ISO, n.d.); Health Data, Technology, and Interoperability (HTI-1) (ONC, 2024b); and The Joint Commission's Responsible Use of Health Data certification (The Joint Commission, n.d.), as well as local governance standards and requirements.

Commitment 3: Engage Impacted Individuals

Health systems and payors should consider prioritizing the engagement of diverse patient groups, health care professionals, and other stakeholders throughout the AI lifecycle to ensure an AI application appropriately addresses patients' needs and respects their values and preferences. Inclusion of ethics and equity experts can help guide the development process, ensuring that an AI solution's primary aim is to advance human health. Continuous feedback from patients, clinicians, and other stakeholders should be gathered and used for ongoing improvements. This feedback should also be shared with patients, health systems, regulators, and accreditors to ensure accountability.

Health systems and payors can *engage with developers* to translate design requirements, goals, and targets into solutions that are adaptable, scalable, and capable of delivering tangible benefits. Partnering with developers and leveraging insights from federal agencies can ensure that AI models are technically proficient, ethically sound, and clinically relevant.

This stakeholder group should *engage workforce and clinician well-being experts and researchers* to collaborate in assessing the impact of AI solutions on clinical workflows and practitioner productivity and morale. In addition, they should *engage AI users* to determine their preferences for how and where the outputs of the AI solutions should be displayed and shared and promote collaboration and alignment with internal technology teams for optimal integration into workflows. Lastly, and most importantly, *patients should be engaged* to identify patient preferences in how, why, and when health AI should be utilized on their behalf, with a focus on goals of care and AI's impact on any decisions made. AI solutions should prioritize patient health, privacy, and well-being (The Light Collective, 2024).

Commitment 4: Improve Workforce Well-Being

As employers of the health care workforce, the health systems and payors stakeholder group have a central role in sustaining and improving the workforce care delivery experience. It is important to prioritize AI technologies that are likely to significantly positively impact the health care delivery experience for the workforce, while also contributing to improved outcomes of care.

Health systems and payors should consider efforts to foster a culture of collaboration and shared purpose among the health care workforce and involve them with agency throughout the AI lifecycle. This includes involving them in decision making that influences the design and development or procurement of AI systems as well as use of regular feedback mechanisms, such as surveys and listening sessions. This should also include prioritizing human-centered AI design and careful workflow integration that fosters communication skills and capacity to preserve and enhance the quality of interactions between health care workers and patients. Organizations should also provide mechanisms and ongoing support for health care workers that includes accepting feedback and addressing moral and ethical concerns as humans learn to adapt and work effectively with AI systems. Effectively establishing this type of culture and transparent processes can facilitate the health care workforce in developing a sense of ownership, investment, and professional satisfaction in the outcomes of AI use.

Workforce training and support in the use of AI in health care delivery is essential for addressing concerns and fostering a positive workplace culture of collaboration and learning, promoting effective and ethical utilization of these technologies. This requires integrating AI training into continuing education offerings, promoting interdisciplinary learning, and establishing continuous professional development programs. To combat automation bias, it is important to enhance AI literacy through targeted and ongoing training on the strengths and limitations of AI. Consideration of decision support systems designed to augment human judgment and highlight potential automation bias for mitigation may be appropriate, as could workflow processes that require explicit review of AI outputs ("AI Timeouts").

Health systems should consider adopting innovative collaboration strategies with health care workforce stakeholders to proactively address the potential displacement of human health care workers and the potential for degradation of human connections due to AI integration (Davenport and Kalakota, 2019). Possible strategies include engaging in open dialogues to discover needs and concerns, conducting workforce impact assessments, and promoting skill diversification toward roles where human expertise remains indispensable.

Commitment 5: Monitor Performance

For AI to be successful in health and health care, its output must be timely and integrated with dedicated infrastructure, resources, and trained personnel. Additionally, AI systems should adapt to changes in the health and health care landscape and focus on meaningful patient outcomes (Kwong et al., 2024), including leveraging higher-quality, real-world data (Silcox et al., 2024).

The efficacy and safety of tools may be compromised over time as evolving clinical environments disrupt the performance of the underlying models (Subbaswamy and Saria, 2020). To maintain the safety and effectiveness of AI algorithms in health care organizations, institutions should establish continuous learning cycles (Embí, 2021), providing ongoing performance monitoring, communication updates to users, and models adjustments or decommissioning as needed. This requires "close collaboration between clinicians, hospital administrators, information technology (IT) professionals, biostatisticians, model developers, and regulatory agencies" (Feng et al., 2022, p. 1). This stakeholder group should engage with appropriate stakeholder groups to establish processes for monitoring ongoing data quality and relevance, which includes transparent reporting of AI performance, including adverse events and success stories (Davis et al., 2020). This also includes the need to evaluate and monitor over-reliance on AI tools in some contexts (automation bias). Overall, promulgating transparency builds trust, ensures accountability, and promotes patient safety and public trust (van Genderen et al., 2024) in AI technologies.

Health systems and payors should consider developing transparent and comprehensive documentation (Brereton et al., 2023) and reporting, outlining how the AI system's performance and impact on health and safety will be monitored and shared. This stakeholder group also plays a central role in promoting accountability, disclosure, transparency, and mitigation of AI-based harms' legal and financial responsibilities in a comprehensive, patient-centered approach. This group also has a responsibility to share the outcomes and impacts of AI applications with other stakeholders, such as vendors, developers, and users, and coordination and facilitating clarity of these shared responsibilities is important for all stakeholders to acknowledge and act on.

Commitment 6: Innovate and Learn

This stakeholder group plays a coordinating role in implementation and maintenance of AI used within health care delivery and should consider adopting a forward-thinking approach to innovation and embrace continuous learning and improvement based on emerging evidence and stakeholder feedback.

Health systems and payors could consider adopting transparent, strategic, and operational alignment of conduct, accountabilities, and relationships (Tabassi, 2023) throughout the AI lifecycle. This alignment is crucial for overcoming siloed approaches and translating ethical principles into actionable practices that drive AI adoption (Adler-Milstein et al., 2022) as a sociotechnical system (McCradden et al., 2023). In addition, defining organizational roles and responsibilities enhances accountability and facilitates effective change management. Stakeholders collectively enhance patient outcomes and operational efficiency through the accountable development, implementation, quality, and risk management of AI in health care (Overgaard et al., 2023).

As health systems and payors collaborate with developers to foster innovation and adapt to the regulatory landscape, it will be important to clarify expectations regarding testing and evaluation before implementing AI into clinical practice could accelerate the benefits experienced by patients. Collectively, stakeholders across the AI lifecycle ensure that AI solutions are ethically developed, effectively implemented, and continuously monitored to enhance patient care and maintain accountability (Bedoya et al., 2022; Raji et al., 2020).

To achieve the goals stated above, this stakeholder group can establish systems and local governance processes for ongoing evaluation, transparency, and communication regarding AI's real-world performance and impact. Organizations can also adopt external standards and develop local standards for what constitutes adequate assessment and learning to know that AI innovations are safe, effective, efficient, patient-centered, and equitable. This ensures continuous improvement and advancement of clinical practice standards (Sendak et al., 2023).

In summary, by considering the AICC Commitments and the AI lifecycle in approaching and using health AI technologies, health systems and payors have an opportunity to optimize clinical outcomes, improve operational efficacy, and foster trust in the health care system. The strategies referenced provide a call to action to develop, implement, and monitor AI solutions in a manner that prioritizes transparency, accountability, and equity. However, realizing these goals requires concerted action from all involved parties.

PATIENTS', FAMILIES', AND THEIR ADVOCATES' PERSPECTIVES

There is a growing awareness that the patient stakeholder voice has been under-represented in health AI (Adus et al., 2023). A recent systematic review of diagnostic AI studies found that patient perspectives were half as likely as clinician perspectives to be represented, although the study was among English-only studies and represents a high-income country bias (Hogg et al., 2023; Kuo et al., 2024). Health AI should be designed with patients recognized as primary stakeholders, end users of AI, and co-creators of responsible AI. It is critical that the patient's voice, needs, and interests shape the future of health AI, in collaboration with other AI ecosystem stakeholders, and patient-led governance and rights are essential for fair, ethical, and equitable integration of health AI in health care. Thus, below are considerations for actions that, as empowered stakeholders, patients and patient advocates could take to contribute to amplify the patient voice in health care AI and embed the Code Commitments into the ecosystem, with alignment to the AI where appropriate.

Commitment 1: Advance Humanity

Patient rights, interests, and unmet needs should be prioritized in the ideation, design, development, deployment, and governance of health AI solutions throughout the AI lifecycle (The Light Collective, 2024). Patients as well as patient advocacy and representation groups can actively engage with other key AI stakeholder groups and continue to advocate for patient stakeholder inclusion and promotion of patient agency throughout all aspects of AI ideation, design, development, implementation, monitoring, and improvement.

Commitment 2: Ensure Equity

Increasingly, institutions, clinicians, government agencies, and the health technology industry involved in developing health AI solutions are recognizing that diverse and under-represented patients and communities bear the greatest burden of risk for health inequities, bias, harm, and discrimination. It is important to advocate for a novel, necessary role in AI governance for patient-led representation, rooted in lived experience and expertise, and relevant to the context of intended AI use.

There are several distinct aspects to such advocacy. Patients and advocacy groups can *assess the degrees of transparency and patient engagement* among organizations using health AI, and advocate for increasing patient representation and equitable use of AI among other AI stakeholders as needed. This stakeholder group can also *champion increased access to their data* for health AI purposes within the context of the cultural values of that group, and the intended uses of the data, with appropriate protections to support expanded AI use that is equitable and representative of that population. In addition to data access, *transparency in and effective communication about what and how patient data are used* in the AI lifecycle for all applications is essential for trust and requires ongoing attention by patients and advocates. It is also important to advocate for transparency of and access to the outputs of AI tools used in a patient's clinical care. Lastly, this group can *advocate and lobby for legal rights and protections* for people and communities negatively impacted by health AI.

Commitment 3: Engage Impacted Individuals

There is an established science to patient engagement partnerships, which can be applied toward fulfilling this Commitment in the context of the AI lifecycle (NAM, 2022b; PCORI, n.d.). AI, by design, should mean co-creation with the patient voice as integral throughout (Vanstone et al., 2023). In the context of health AI, this includes *promotion of mechanisms, methods, outreach, and frameworks to lower barriers to engage patients* in each stage of the AI lifecycle. Patients and advocates might also work to promote awareness of patient sub-populations (e.g., those with specific diseases or cultural values and experiences) whose perspectives might be particularly suited to specific AI solutions.

Patients and advocacy groups have an opportunity to support the expansion of funding and investments in research that requires their engagement. They can advocate for research funding announcements to require patient participation in health AI research project design and oversight, and they can engage professional societies and present at research conferences to promote inclusion of patient perspectives at AI-related venues.

The advancement of explicit inclusion and representation for patients and advocacy groups in the governance and processes of AI use among health systems and payors is essential. Establishing standing patient representation groups collaborating with one or more organizations along specific patient missions related to disease management or cultural values can facilitate continued engagement.

Commitment 4: Improve Workforce Well-Being

As the health care workforce's mission is to care for patients, this stakeholder group can play a central role in partnering with health care workers to improve the quality, safety, human connection, and trust in the delivery of health care. By advocating for health AI that meets patient needs, perspectives, and cultural values while simultaneously promoting outcomes that improve the workforce's sense of purpose and capacity for patient engagement, patients and advocates may gain a greater sense of partnership with their clinical teams. For example, this group could advocate for AI that improves the time health care workers can spend in direct interaction, as the most common reason for patients to lose trust in their provider was spending too little time with them (Birkhäuer, 2017; NORC, 2021).

Commitment 5: Monitor Performance

Trust in health AI will be aided by transparency in monitoring of AI technologies' performance related to safety, efficacy, and equity that is comprehensible to patient communities and advocates. Patients and advocates could consider calling for concepts, frameworks, methods, and tools that can translate the technical and clinical specifications and needs of health AI into patient-centered, culturally appropriate language and interpretation. More specifically, examples could be to advocate for disclosures in accessible terminology similar to nutrition labels on foods (Sarasohn-Kahn, 2021), transparency of bias assessments, applicability of AI tools to the populations they are being used in, and safety and adverse event reporting in the context of use. If AI-generated clinical decision support outputs are used to make decisions about an individual's care or coordination of care, this group can advocate for individuals having a protected right of access to these outputs, and request that said outputs be documented in the individual's patient record. They can also advocate for such outputs being recognized as a part of a patient's designated record set, recognized as electronic health information (EHI) (HHS, 2024c), and for having information blocking rules (ASTP ONC, 2024) applied (The Light Collective, 2024).

Commitment 6: Innovate and Learn

Given the novel issues that AI presents, particularly for patients who provide the data to train AI models (monetized by others) and who are recipients of AI outputs, sometimes without their explicit knowledge, innovation in governance is an important consideration to ensure that the primary aim of advancing humanity through health AI is achieved.

Patients and advocates can champion a requirement for patient representation that holds legally enforceable an independent duty of loyalty to continuously improve outcomes for patients. They can advance the imperative that this oversight have real authority, especially when industry or health system interests conflict with the public good, patient safety, and privacy.

This group can advocate that resources and funding be prioritized and allocated for diverse patient communities to conduct outreach education to build capacity for patient representatives to be informed consumers of AI and facilitate capacity for patient organizations to establish and promote independent oversight. This investment could reinforce the concept that patient voices are integral to the development and deployment of health AI technologies in a sustainable and longitudinal manner.

Due to the transformative nature of AI, some aspects of individual consent may no longer be adequate to protect the health, safety, and well-being of patients. This group could call for stakeholders to proactively co-develop consent and community governance frameworks for health AI use.

It is important for patients and their care partners to understand what AI-powered clinical decision support (CDS) tools and innovations that different health care delivery systems and providers use. In the same way that patients may search a health system's directory of physicians to explore their biographies and clinical interests, it would be useful for patients to be able to access a directory of CDS tools that may be utilized at a given health system. Patients may also want to see which physicians may incorporate these tools into their care. This level of openness can build trust and may invite shared decision making. It may also create opportunities to advance patient agency.

Advocacy could also include the right to opt out of the use of AI-powered technologies, especially if it increases the costs of their care without a significant impact on outcomes, the disclosure of what data were used to train specific AI-powered CDS tools that may be used in their care, and how their data are being used in relation to AI.

The American Hospital Association's Patient Care Partnership (formerly the Patient Bill of Rights) (AHA, 2003) provides an overview of what patients should expect during their hospital stay. Patients and patient advocacy groups could advocate for an update to this document to reflect new inclusions that relate to AI. This might include informing patients about the role of AI in their care if AI-powered CDS tools are used to guide diagnosis and treatment, as well as the potential benefits, risks, and limitations.

FEDERAL AGENCIES' PERSPECTIVE

In 2024, the Department of Health and Human Services (HHS) reorganized technology, cybersecurity, data, AI, and policy functions under the Assistant Secretary for Technology Policy Office of the National Coordinator for Health Information Technology (ASPT ONC) and has appointed a chief AI officer and a chief data officer, to report to the chief technology officer (HHS, 2024b). And HHS published an AI Strategy, a Trusted AI Playbook, and its AI Strategic Plan (HHS, 2021, 2025, n.d.a). The Agency for Healthcare Research and Quality

(AHRQ) Digital Health Care Research Division reported in 2022 that 17% of its grants since 2020 involved health care AI research (AHRQ, 2022). CDC is funding partnerships with private-sector experts to develop a framework for identifying and preventing biases in public health uses of AI tools (CDC, 2022). More activity related to health AI is expected in the coming months in the wake of these changes.

Federal agencies could use these same authorities and tools to promote alignment with the AICC Commitments. This requires a recognition of the different statutory missions and authorities of each agency as well as a commitment to shared strategic and operational alignment to ensure a consistent message to developers, patients, caregivers and others who will be affected by the creation of new AI tools for health. Within that context, outlined below are considerations of ways in which federal agencies could incorporate the AICC Commitments in their internal research, development and clinical care efforts, and in their external stakeholder engagement, standard setting, regulatory activities, and program priorities for funding. Where applicable, alignment of these activities with the AI are noted.

Commitment 1: Advance Humanity

This commitment directly aligns with HHS's mission "to enhance the health and well-being of all Americans, by providing for effective health and human services and by fostering sound, sustained advances in the sciences underlying medicine, public health, and social services" (HHS, n.d.b). This concept is already central to all federal agencies involved in human health, including those outside of HHS. However, in the context of the challenges for health AI, federal agencies play a central role in providing appropriate protections, regulations, and policies to ensure continued human agency in health and health care.

- The National Science Foundation, CDC, AHRQ, U.S. Food and Drug Administration (FDA), the National Institutes of Health (NIH), the Health Resources and Services Administration (HRSA), and other federal agencies could fund additional studies to measure how AI can influence patient health, with particular emphasis on how AI influences human agency, goals of care, and human–human interactions in the presence of AI interventions.

Commitment 2: Ensure Equity

As noted above, federal agencies have existing policies, recommendations, and guidance to address this commitment. However, large gaps remain in effectively ensuring equitable benefits and risks to all, across technical, sociotechnical, and cultural considerations and settings. This field is evolving rapidly, and agencies can play a pivotal role in shaping how bias and equity in health care are addressed through research investments as well as iterative policy and standards updating. ASTP ONC can support coordinated efforts to fill these gaps.

- NIH, the Centers for Medicare & Medicaid Services (CMS), CDC, and other agencies could fund additional research that uses AI to measure disparities; agencies could also further prioritize funding to evaluate the extent to which AI interventions reduce or increase disparities. All federal agencies that leverage AI prediction models could require representative sub-populations in models developed and used, as well as support expansion of methods and tools to comply with requirements for evaluation and mitigation of bias and inequitable performance in sub-populations.
- As assessment and mitigation of bias mature, FDA, ASTP ONC, and other regulatory agencies could iteratively expand requirements in the use of diverse and representative training and validation datasets to help improve performance across sub-populations.
- Federal agencies could continue to update data sharing requirements and develop and expand sustainable data sharing platforms such as HHS's HealthData.gov to provide health AI application development with improved access to diverse and representative data.
- In undertaking enforcement, the ASTP ONC and the HHS Office for Civil Rights (OCR) could prioritize information blocking enforcement and HIPAA enforcement to enable individuals' access to EHI, including EHI that is created by AI, to support openness and engagement of individuals. OCR has promulgated regulations (45 C.F.R. Part 92 implementing Section 1557 of the Affordable Care Act) that prohibit discrimination from patient care decision support tools. OCR could build on these regulations to support transparency regarding data sources.
- ASTP ONC could recognize standards for collection and exchange of relevant data and encourage use of the Trusted Exchange Framework and Common Agreement (TEFCA) for making data available for prioritized research projects, including using the individual access service. Adoption of the commitments could be required for researchers seeking access via TEFCA for such priority research data sharing (ASTP ONC, 2024).
- CMS could also enable compliance with the Code Commitments to protect and advance human health by prohibiting the use by health plans under its purview of AI-powered claims denial processes in ways that harm patients.

CMS has already taken a significant step in this direction by requiring Medicare Advantage plans to make coverage decisions based on individual circumstances versus relying on generalized algorithms.
- OCR could expand guidance on how AI tools can be leveraged, consistent with HIPAA and applicable civil rights laws, such as Section 1557, and provide guidance on how to monitor for discriminatory impact of patient care decision support tools.

Commitment 3: Engage Impacted Individuals

- Research funding agencies could undertake and fund research and evaluation regarding stakeholder engagement in AI, evidenced by engagement records, diversity of input, and stakeholder satisfaction of research products.
- Federal agencies could expand stakeholder engagement to ensure that agency-wide funding opportunity announcements (FOAs), research funding decisions, regulatory approval processes and health care delivery standards include evolving concepts of inclusivity, equity, and representativeness in the context of use.

Commitment 4: Improve Workforce Well-Being

- CMS could monitor and fund others to assess the well-being of the health care workforce interacting with AI systems, using surveys, and assessing changes in job satisfaction and perceived workload.
- CDC could continue to expand evaluation of the state of the public health workforce through development, integration, and assessment of new AI tools that identify potential ways to enhance the workforce well-being.

Commitment 5: Monitor Performance

- NIH could collect and make public information on AI system performance, including access to performance data, adherence to regulatory reporting requirements, and active dissemination of findings to all stakeholders. This could be achieved through a system such as ClinicalTrials.gov, which is supported by the National Library of Medicine and FDA.
- ASTP ONC has already taken action to require transparency and risk management for AI-assisted decision support interventions that are part of certified EHR products (HHS, 2023a); ASTP ONC could evaluate and consider the Code Principles and Code Commitments for inclusion in regulations and guidance regarding conditions and maintenance of certification.

- The Veterans Health Administration (VHA) could promote this commitment through open sharing of data, algorithms, performance, and metrics of AI solutions implemented within VHA.
- The Cybersecurity and Infrastructure Security Agency could use its authorities and influence to promote additional transparency and standardized reporting requirements for security and privacy assessments of AI tools before, during, and after deployment.

Commitment 6: Innovate and Learn

As part of the overall ecosystem of safe and effective innovation and continuous improvement, federal agencies could incorporate the Code Principles and Code Commitments, where not already present, in FOAs for research (Public Health Service agencies such as NIH and CDC), criteria for product approvals (FDA), and conditions of participation in Medicare and Medicaid (CMS).

- NIH, CMS, FDA, ASTP ONC, and other federal agencies could expand funding for research into the rate and impact of continuous improvements to AI systems, evidenced by update logs, the implementation of feedback into system refinements, and innovation metrics.
- HRSA, the Indian Health Service, VHA, and the Department of Defense could expand research into the development and implementation of AI-supported CDS tools integrated into health IT systems that address inclusion and equity.
- FDA has already issued an action plan for AI-based, medical device software (FDA, 2021) and put forth discussion papers on AI in drug development (FDA, n.d.). The agency could propose or require additional levels of adherence with the collection of real-world evidence that adheres to broad diversity and equity considerations (FDA, 2024).
- CMS has already established payment methodologies for reimbursement of AI products (Chen et al., 2021); the agency could further adopt reimbursement standards that prioritize compliance with the Code Commitments and products that are built on or implement the commitments.
- ASTP ONC could expand certification standards or conditions of certification in health IT technologies, including but not limited to EHRs, regarding to continuous learning and updating AI over time.
- As the Health Systems & Payor workgroup observed, collaboration between researchers and federal agencies can set the standards for necessary oversight and support for adapting AI solutions to emerging challenges and shifting patient needs. HRSA and CMS both have authorities to lead on this work.

- FDA could expand requirements in AI product labeling based on real-world performance and other characteristics that would assist users and patients in understanding the context and performance of the use of the tools.

HEALTH AND HEALTH CARE WORKFORCE PERSPECTIVE

The well-being of the health and health care workforce has faced numerous challenges over the years from various occupational stressors, such as a historical culture of professional focus that challenged work/life balance, usability issues and time spent on the EHR, and most recently the COVID-19 pandemic (Alobayli et al., 2023; Shanafelt, 2021). In addition, well-being and global health care delivery are threatened by a growing shortage of health care personnel due to several factors, with the World Health Organization estimating that by 2030 there will be a shortage of 14 million nurses, physicians, midwives, and other health care professionals (deVries et al., 2023; WHO, 2016). In the setting of this growing crisis and amidst a health care culture adopting an increasingly proactive framework of systems-based interventions to identify and address causes of occupational distress (NAM, 2019), AI has substantial potential to harm or renew the well-being of the health care workforce. Presented below are key opportunities and challenges in the use of AI in health care in the context of the Commitments through the lens of Commitment 4: Improve Workforce Well-Being.

In addition, health systems and payors should leverage lessons learned from EHR implementation, which is discussed in another context in Chapter 2. EHR use, promoted by the HITECH Act, has been associated with decreased incidence of adverse events and occasionally improved coordination of team-based care and chronic disease management, but also decreased job satisfaction and higher burnout rates (Bates et al., 1998; O'Malley et al., 2015; Peccoralo et al., 2021; Reed et al., 2012). For EHRs deployed in environments with heavy regulatory compliance requirements and documentation burdens to satisfy multipayor reimbursement paradigms, perceptions that EHRs facilitate billing rather than clinical care are unsurprising (Holmgren et al., 2021; Tseng et al., 2018). Such challenges, which were anticipated by a 2009 National Research Council report (NRC, 2009), elucidate opportunities to leverage EHRs in reducing cognitive load and increasing cognitive support (Johnson and Stead, 2022). These opportunities are germane to AI tools in health care, which have the potential to reduce cognitive burden at the price of imparting automation bias in which health care workers experience a loss of critical thinking, judgment, and intuition—elements that must be preserved in AI-enabled decision support design and deployment (IOM, 2003).

Commitment 1: Advance Humanity

Integrating AI into health care workflows is a challenging task, and could involve functions such as automation, cognitive support, or information synthesis, among other opportunities. However, health care AI applications could degrade human connections among health care workers and between providers and their patients by automating tasks that have, historically, involved human-to-human interaction (e.g., a clinician obtaining sensitive information from a patient or their caregivers). Conversely, automating some basic tasks (e.g., data capture, medical coding) could reserve time and space for health care workers to be with one another and patients. Another challenge is that if simple, straightforward tasks are automated, humans could face workdays exclusively comprising complex, demanding tasks (e.g., a computer vision model diagnosing unremarkable and low-acuity chest radiographs while triaging images with moderate or severe pathology to human radiologists, who see nothing but disease, injury, and diagnostic challenges) (NEJM-AI Grand Rounds, 2023). Balancing the needs and preferences of all stakeholders while simultaneously prioritizing human health is essential to ensure that AI advances human agency and connection. User-centered and equitable approaches to the design and deployment of AI tools are best practices and are required to ensure they meet the varied needs of patients and health care providers effectively (Van den Bruel et al., 2010, 2012).

Commitment 2: Ensure Equity

The health workforce has a critical role in addressing disparities in health care service delivery, and workforce diversity is foundational for health equity (Pittman et al., 2021). Occupational stressors can be felt differentially among constituencies in the health care workforce, as was found in higher rates of unemployment among those self-identifying as Black or Indigenous following the COVID-19 pandemic (Semprini, 2023). AI technologies may be implemented in ways that are not culturally aligned with practice preferences for under-represented health care workers, or not aligned with patient preferences, and exacerbate rather than ameliorate workforce well-being and reduce equitable care (Parag et al., 2023). Care should be taken to understand the context of use for AI technologies and anticipate their use in under-represented populations and environments (e.g., rural practices and community health centers).

Commitment 3: Engage Impacted Individuals

Deploying AI in health care settings risks displacing health care workers, such as when AI replaces human work products (e.g., AI performing diagnostic tasks on radiographic and pathologic images). In the automotive industry, when robotic arms replaced assembly line workers for basic tasks, automobile prices fell while human workers experienced job loss. Eventually, jobs shifted toward tasks requiring creativity, long-term planning, and moral deliberation, skills that remain relevant for recent automotive industry endeavors, such as self-driving cars (Awad et al., 2018).

Given the high probability of similar job shifts in health care, it will be important to identify training and employment needs in areas less likely to experience AI automation, and training and retraining humans to work with AI. Education-enabled health care workforce job shifts, if successful, could not only mitigate risk of job loss, but also address global health care workforce shortages by maintaining current workforce volumes while addressing occupational stressors through workflow and work product optimizations with AI tools. For example, while not yet ready for widespread use, AI-enabled medical interpretation services may at some point replace human medical interpreters, easing a serious resource shortage and improving patient and provider experience (Lion et al., 2024); retraining those workers for roles that require more creativity, planning, and moral deliberation could benefit the workforce and the system.

Commitment 5: Monitor Performance

One of the challenges in health care workforce well-being has been the measurement and assessment of well-being in complex settings. Historically, most validated instruments of well-being assessment were manual and not easily monitored in an automated fashion (Boskma et al., 2023). The use of AI to predict well-being for downstream management is a growing field (Levin et al., 2024; Nan et al., 2024) with opportunities to identify policies, workflows, environments, and circumstances that negatively impact the ability to improve workforce well-being (Malgaroli et al., 2023). However, if these tools are not appropriately used, they could directly harm health care workers and lower well-being. As noted elsewhere, any AI tools put into place for these types of activities themselves should be monitored for performance and appropriateness in the context of use.

In addition, from a legal perspective, health care workers who use AI tools may incur personal liability for AI-associated errors or patient harm (Mello and Guha, 2024). As tort doctrine evolves to address the distinctive challenges posed by AI, health care workers must be informed of the risk for personal liability and relevant institutional policies and should be protected by surveillance of AI-associated outcomes that adapts to the probability and severity of adverse events (i.e., higher-risk AI tools will require heavier surveillance). Those who purchase AI tools may have contracting opportunities to shift liability to developers when model outputs are erroneous and cause harm, while the purchaser or health care worker would remain liable for AI tool misuse (Banja et al., 2022). could also protect health care workers and patients by insisting that AI tools are demonstrably safe, effective, and meeting salient (e.g., FDA Software as Medical Device) standards (Stern et al., 2022).

Commitment 6: Innovate and Learn

Effects of AI on the health care workforce will likely include elements that are currently unknown and cannot be foreseen. Mitigating risk for "unknown unknown" or "black swan" events should focus not on prediction—which is impossible—but rather should identify areas of vulnerability and build robustness against them (Rumsfeld, 2011; Taleb, 2007). LHS principles that support the translation of data to knowledge to practice are applicable here. The goal is to ensure that clinical and operational knowledge gained from AI development and implementation is incorporated in AI applications that affect care delivery in a manner that compounds as the health of patients, health care workers, and health care system evolve (IOM, 2011). This approach may facilitate early detection and effective rethinking and adaptation when impactful, unforeseen events occur. Like all other strategies and tactics pertaining to health care workforce considerations, this approach must involve and represent all stakeholders in shifting the influence of AI away from harm and toward the well-being of health care workers.

Similarly, when some types of AI applications become reimbursable, there is risk for incentive-driven overuse and increasing documentation burdens for billing purposes. These risks may be mitigated via value-based incentives informed by audits of patient- and provider-level associations among AI use, coding practices, and patient outcomes.

HEALTH CARE QUALITY AND PATIENT SAFETY PERSPECTIVE

In 1998, the Institute of Medicine (IOM) Roundtable on Health Care Quality described an urgent need to improve the quality of health care, defined as "the degree to which health care services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge" (Chassin and Galvin, 1998). Harms related to overuse, underuse, and misuse were identified and further described in patient safety literature (Emanuel et al., 2008). Despite significant efforts, improvement and comprehensive measurement of health care quality and patient safety have proven challenging. In their article, "Two Decades Since *To Err Is Human*: An Assessment of Progress and Emerging Priorities in Patient Safety," the authors noted that "the frequency of preventable harm remains high, and new scientific and policy approaches to address both prior and emerging risk areas are imperative" (Bates and Singh, 2018). Additionally, measurement is critical to improving quality, and while there is considerable dissatisfaction with the current state of measurement and assessment, few serious efforts exist to make significant changes (McGlynn, 2020).

There are "islands of excellence" as is demonstrated by the reduction in adverse events associated with major surgical procedures, cardiac events, and pneumonia between 2010 and 2019 (Eldridge et al., 2022). However, in 2022, the U.S. Office of Inspector General reported that 25% of Medicare patients experienced harm during their hospital stays during its study period (OIG HHS, 2022). Post-pandemic, U.S. health care safety declined precipitously and severely (Fleisher et al., 2022). Ensuring that the health care system heals rather than harms patients remains a critical priority for the United States; however, routine, comprehensive, and systematic national assessments of safety have not been established. In fact, the last national assessment was conducted in 2003 (McGlynn, 2020).

The Role of AI in Improving Quality

As described in previous sections of this chapter, AI has significant potential to enhance the quality of health care through improved prevention, early detection, diagnostics, and early intervention, treatments, and rehabilitation. AI may also be a contributing factor in catalyzing a new era of quality improvement through enhancing the accuracy, efficiency, and comprehensiveness of quality measurement, while making it substantially less burdensome. AI's capabilities to process vast amounts of data rapidly and accurately contribute to enhanced quality measurement in health care (Jiang et al., 2017).

Quality measurement begins with data collection and integration. AI technologies can streamline data collection from various sources including EHRs, social determinants of health databases, wearable devices, and patient self-reports. By automating the extraction and integration of data, AI reduces the burden on health care professionals and ensures more comprehensive and accurate datasets (Esteva et al., 2019). This comprehensive data collection enables a more detailed and nuanced understanding of patient outcomes and care quality. Additionally, AI has the ability to process and evaluate vast datasets, identifying relationships and repeated patterns that might otherwise go unnoticed by humans or be undetected using traditional statistical techniques (Shickel et al., 2018).

Traditional quality measurement often relies on standardized metrics that may not fully capture the unique needs and circumstances of individual patients. AI can create personalized quality metrics by including information about the individual, such as specific health conditions, treatment preferences, and social determinants of health (Obermeyer and Emanuel, 2016). Personalized metrics provide a more accurate reflection of care quality and patient satisfaction, in that AI can improve the measurement of patient outcomes by providing more sophisticated tools for evaluating treatment effectiveness and patient satisfaction. For instance, AI-driven sentiment analysis can assess patient feedback from surveys and social media, offering insights into patient experiences and areas for improvement (Topol, 2019). AI can also track long-term outcomes more effectively, ensuring that quality assessments reflect the sustained impact of care interventions. By analyzing data across multiple health care institutions, AI can identify best practices and facilitate benchmarking. Comparative analyses can highlight high-performing institutions and effective care strategies, guiding quality improvement initiatives (Rajkomar et al., 2019).

The Role of AI in Improving Patient Safety

Multiple interconnected factors influence patient safety. For example, harms can result from challenges with data, interoperability, system usability, clinician burden, and complexity of clinical scenarios (Tighe et al., 2024). AI, with its capacity to incorporate vast and disparate datasets, represents an important tool that holds vast potential to improve patient safety. AI has been successfully tested across eight major harm domains, and situations where new or unstructured data sources are incorporated to improve predictions are expected to yield the highest impact on patient safety. These novel data include sensing technologies such as "vital sign monitoring, wearables, pressure sensors, and computer vision" (Bates et al., 2021).

Using these data sources, AI models are poised to significantly improve risk prediction and thereby advance patient safety. Integrating data across novel and disparate sources in real time and adapting to continuously updated data, AI models can deliver timelier and more accurate predictions based on information such as biometric, sensor data, or video, sources that are currently untapped or abstruse (Tighe et al., 2024). An example of this is the use of movement trackers and camera data to identify fall risk (Tighe et al., 2024).

AI and Patient Safety and Quality in the Context of the Code Commitments

While the potential for improving health care quality and patient safety with AI is high, it also comes with opportunities and risks beyond those discussed previously. The Code Commitments offer additional guidance to mitigate these risks.

Commitment 1: Advance Humanity

As previously described, AI has the potential to significantly improve both human health and well-being. This must remain the litmus test for all applications of AI in health care. Generative AI can also create high-quality human-like responses to patient inquiries (Ayers et al., 2023); however, such responses—agency enhancing and available free online to individuals—can produce information that is inaccurate and potentially harmful (Coiera et al., 2023). It will be imperative to educate the public about the benefits and risks of AI that appears to provide empathetic medical advice outside the confines of trained health professionals.

Commitment 2: Ensure Equity

Equity is not limited to the outputs of AI but also to the access to the benefits of AI. If AI systems that improve quality and reduce harm are implemented, ensuring that those systems are available to low-resourced health systems will be essential for true equity. AI can be particularly susceptible to inequitable performance from biased training data and misaligned design but also can be designed explicitly to mitigate bias and promote equity (Cary et al., 2023) if conducted in a systematic way. Consideration of financial incentives for implementation of proven AI systems by small community and rural hospitals is warranted.

Commitment 3: Engage Impacted Individuals

The potential of AI to cause harm if poorly designed and implemented is as significant as with EHR implementation which contributed to an increase in errors and patient safety concerns due to information overload, presentation of irrelevant information, and data display issues (Nijor et al., 2022). Even more complex, when AI is integrated into EHR-based CDS tools, then potential safety challenges may occur from either or both in synergy and may still contribute to the already substantial problem of alert fatigue. Lack of user engagement in system design and implementation of new workflows can result in mistrust among patients and clinicians, diminishing the anticipated gain from the use of AI

(Tighe et al., 2024). Ensuring collaboration with clinical and administrative teams in the design and implementation of AI systems will be essential.

Commitment 4: Improve Workforce Well-Being

The application of AI in health care presents opportunities to reduce burden for clinicians; however, it also presents risks of over-reliance on AI, clinical de-skilling, and ineffective human oversight. Relying on manual review of AI-generated decisions or recommendations is a weak approach to ensuring patient safety (Lucian Leape Institute, 2024). It is highly likely that as AI tools become more prevalent, clinicians will not be able to consistently review the output of AI scribes, bots, and other tools to spot errors or carefully reflect on the diagnostic or therapeutic suggestions of decision support tools. Thus, the error rate or risk of harm from any given AI tool needs to be considered on its own, not presented as if double-checking by clinicians will prevent those errors from impacting patients (Lucian Leap Institute, 2024).

Additionally, health care organizations, seeking to defray the costs of acquiring AI tools, could use AI-driven efficiencies to assign more duties to clinicians, nullifying anticipated improvements in workloads and cognitive burdens and, combined with de-skilling, further reduce professional satisfaction. AI will not produce the anticipated benefits in health care if the underlying health care payment model, which continues to pay for quantity over quality, is not simultaneously addressed.

A recent survey by the National Association for Healthcare Quality noted that the health care quality and safety workforce often have a low level of competency in health data analytics and performance and process improvement (NAHQ, 2022). It is critical to facilitate and promote a quality and safety culture that has the training to adapt to an AI-enabled health environment.

Commitment 5: Monitor Performance

To proactively monitor and mitigate risks to patient safety, AI models and tools require continuous monitoring after implementation (Feng et al., 2022). AI models can be designed to monitor other AI models and tools to detect drift in the underlying data, performance loss over time, and differential performance in sub-populations (Davis et al., 2017). It may be advisable to consider the creation of a specific area of focus within health systems accountable for quality management and continuous improvement of AI models and tools, which have been referred to as "AI-QI" units (Feng et al., 2022).

Commitment 6: Innovate and Learn

Key to quality improvement and patient safety is seeking to continuously learn and improve care based on data and experience. AI offers immense opportunities to advance quality and safety in an LHS. Prioritizing research in this arena is essential to creating a strong evidence base for improving health outcomes. In addition, ensuring collaborative knowledge sharing across health systems will be important to rapid system-wide learning and diffusion of innovation.

AI is poised to support major advances in patient safety and quality measurement and improvement in health care. Efforts to leverage these opportunities must be balanced by those to mitigate the risks in an evolving landscape, aligning with the Code Principles and Code Commitments to ensure safe, high-quality, and trustworthy AI in an LHS.

BIOETHICS AND EQUITY PERSPECTIVES

The objective of this section is to highlight ethical and equity aspects related to the Code Principles and Code Commitments that may be in potential conflict as well as some examples for how to resolve such potential conflicts. Here, core principles and concepts from bioethics are brought together with attention to equity, conceptualized as the social and structural factors that affect individuals' and populations' abilities to achieve health.

There are a number of relevant ethical conceptual frameworks applied in scrutinizing ethics in AI (Heilinger, 2022; Zong and Matias, 2022). For this section, the Code Principles are conceptualized as *substantive* principles, meaning that they help guide how AI should be implemented in terms of characteristics that directly impact human health. The Code Commitments are viewed as *procedural*, meaning that they should help guide the process for how AI tools are developed, and help guide balancing and deliberations when substantive or other principles or values come into conflict.

From an ethics and equity perspective, it is important to go beyond the list of principles and commitments and consider how these may be in tension with one another; how—in case of tensions—they may be weighted or how this tension should be considered, negotiated, and even resolved; and even how they may complement each other. Detailed below are examples of such tensions to illustrate the importance of this perspective.

Safe and Effective May Come into Tension with Transparent

Emphasizing or requiring opening of the black box may reduce the safety or effectiveness of an AI tool in health. This may occur as a function of cognitive reliance on the AI tool without the ability to assess underlying information (Jabbour et al., 2023), which may result in reduced safety. It may also occur if understanding the data context or patient context supporting the AI could result in improved shared decision making (effectiveness).

Adaptive May Come into Tension with Accountable

Adaptive involves continuous learning and improvement, which are laudable goals, but ongoing change and evolution may make it challenging to maintain "clear accountability for potentially adverse consequences." Put simply, change may cause ambiguity regarding who is accountable for AI-caused harms.

Equitable May Come into Tension with Secure

Prioritizing equitable access to data, for example, can compromise considerations related to protection of privacy. This may occur in the setting where patient anonymity cannot be assured for data contributed by under-represented patients (Brown et al., 2023). Inclusions of data from these patients increases privacy risk, not including data from these patients decreases equity in AI development and performance.

Engaged May Come into Tension with Efficient

Engaged involves prioritizing the needs, preferences, and goals of people, which may not always align with the goals of reducing costs for health gained or provide the most rapid mechanism to implement AI.

The Code Commitments may assist in identifying ways to resolve such tensions or recognize complementarity between principles. For example, "involvement" tells us to always engage people as partners. This means that some conflicts may be resolved by discussing with the relevant partners what matters to them in specific contexts. The use cases that follow provide examples of how the Code Principles may come into tension, and how turning to the Code Commitments as procedural tools may offer some paths forward.

Illustrative Scenario 1: Tension Between Safety and Transparency: Documentation Support

Clinicians often grapple with burnout associated with administrative burden of EHR documentation such as completion of clinical notes to ensure accurate billing and addressing inbox messages from patients, care team members, and consultants (Budd, 2023; Tran et al., 2019). Generative AI may be able to automate much of this administrative work and allow clinicians to focus on patient care (Reddy, 2024). Studies show AI-assisted documentation reduces burdens and frees clinicians to better serve patients (Tierney et al., 2024). AI has also shown the capacity to resolve inaccuracies in EHRs, such as discrepancies in medications, thus enhancing patient safety (Damiani et al., 2023).

However, the use of such AI tools raises concerns that patients may not understand how AI is being used in their care, potentially compromising their autonomy. Even if patients have the option to consent to—or decline—the use of AI, they may not fully comprehend what they are consenting to or declining. Moreover, if AI support in documentation is demonstrated to improve patient safety in certain use cases and becomes standard of care, equity challenges can arise if patients decline the use of these tools, especially if the decliners disproportionately come from marginalized populations. Hence, safety and transparency may come into conflict, raising ethics and equity issues.

One possible approach to such conflicts would be, as suggested above, to rely on the Code Commitments for guidance. Focusing on human health and connection, ensuring equitable distribution of benefits, involving people as partners, and promoting transparent stakeholder-inclusive prioritization between conflicts are all procedural requirements that can help negotiate such tensions in specific use cases.

Illustrative Scenario 2: Tension Between Adaptability and Accountability: Automated Diagnosis

IDx-DR was the first FDA-approved autonomous medical AI device. This tool diagnoses an eye disease, diabetic retinopathy (DR), by analyzing images of the eye. This AI device is significant because it provides a *diagnosis*, rather than *advice* to a clinician about a disease or condition (Van der Heijden et al., 2018). IDx-DR can provide a diagnosis and can be used by non-specialist clinicians, whereas previously this condition would have to be diagnosed by a specialist. Because of AI-assisted DR detection tools, DR can be diagnosed by a primary care physician. There is already promising evidence that these tools are allowing more people to be screened for this disease, which can lead to improved health outcomes, such as increases in patients receiving eye screening exams (Knapp et al., 2023). There are increasing numbers of studies evaluating the efficacy and utility of automated DR detection (ADRD) (Joseph et al., 2024).

ADRD systems are powered by machine learning, which means that their diagnostic capabilities are derived from training data. In this way, these are adaptive tools, which is in line with the Code Principle of adaptability. Although adaptability is important for its primary diagnostic function, this adaptability can be in tension with the principle of accountability. Clinician users and patients may not be aware of changes in performance or limitations in environment or setting, and it might be difficult for clinicians and developers to account for or explain the source of errors. In this case, when adaptability and accountability can be in conflict, it is important to turn to the Code Commitments to balance proposals for actions and next steps.

In this case, both the adaptability and the accountability of the ADRD tools advance human health. In these situations, prioritization of one commitment over another needs to be done, and this should be conducted in the context of use, with engagement from all relevant stakeholders. In the setting of this section focused on ethics and equity, a prioritization of accountability promotes equity and prompts reflection on whether the errors might be disproportionately occurring within certain populations. There are other situations in which an AI tool is used in life-threatening clinical situations in which highly dynamic pre-specified change control policies need to be in place to ensure safe operation. In this case, adaptability of the tool for safety reasons may outweigh accountability considerations.

Illustrative Scenario 3: Tension Between Security and Equity: Secondary Use of Data and Consent

There has been a long-standing tension in the secondary use of routinely collected health data between opt-in and opt-out consent procedures (de Man et al., 2023; Sanderson et al., 2017). While there have been variations in individual studies, a majority of patients interviewed or surveyed expressed both a desire for control over their data and a willingness to participate in data sharing for promotion of human health and particularly the health conditions of interest to the patient, and with less interest in those applications with commercially profitable objectives (Kalkman et al., 2022; Mikkelsen et al., 2023; Skovgaard et al., 2019). In health AI, this can be challenging as these goals are pursued simultaneously in many applications.

Perhaps counterintuitively, opt-in models—which appear to improve security and privacy and individual control of health data—may result in bias in consented datasets. A recent meta-analysis of opt-in versus opt-out studies found that the average weighted consent rate was 84% for opt-in and 96.8% for opt-out; however, when both procedures were explained "the consent rate was 21% in the opt-in group and 95.6% in the opt-out group" (de Man et al., 2023). In the context of data use for health AI development, this becomes particularly challenging due to the volume of data necessary to rigorously train these algorithms. Additionally, across the studies, the opt-in models yielded more biased datasets, with represented individuals more likely to be male, have more education, higher earnings, and improved overall socioeconomic position (de Man et al., 2023). While bias in health AI can come from several sources, bias in source data used for training is one of the most established and studied (Ferrara, 2023). Among these, selection bias can result in lower performance and disadvantage among under-represented populations in the data (Haneuse, 2016; Johnson et al., 2000).

In this case, the security and equity associated with secondary data use can appear to be in conflict, but the commitment to ensure equitable risks and benefits prompts reflection about the benefits of secondary use of health data for AI being disproportionately provided to some sub-populations and withheld from others. Additionally, the commitment to transparently monitor performance, including for evidence of bias, both proactively and retrospectively, can reduce the risk of disproportionate benefits or risks to any subgroup.

Considering the Code Principles and Code Commitments together to address apparent tensions is a valuable construct for researchers, developers, patients, and other stakeholders who can use them as a guidepost for designing, developing, implementing, monitoring and maintaining AI tools in health care.

COMMON THEMES AND CONTRIBUTIONS ACROSS STAKEHOLDERS' PERSPECTIVES

The expert working groups, representing broad stakeholder constituency, considered important actions to disaggregate and advance the Code Commitments in the context of the AI. Both similarities and differences were identified between and among the outputs of the working groups. Summarizing these commonalities and distinctions, Table 5-2 identifies needed actions repeatedly identified across groups, while Table 5-3 describes the role each stakeholder group can play in applying the Code Principles and Code Commitments.

TABLE 5-2 | Common Themes for Action Among Expert Working Groups

Commonalities in Stakeholder Actions in Translating the AICC into Action
Ensure that patients, end users, and ethicists are represented throughout the entire AI lifecycle
Ensure that the utility and effectiveness of health AI tools are initially and continuously assessed and optimized for both technical and health outcomes
Promote transparency and documentation of the characteristics, capacities, data sources, intended uses, and limitations of health AI applications
Promote a continuous learning environment in the context of use of health AI applications, with elements of education, iterative improvements, and establishing a culture of systems-based learning and quality improvement
Recognize that conflicts of interest and stakeholder objectives will occur and develop a process of prioritization that considers established ethical frameworks and the AICC Code Principles and Code Commitments for resolution of these issues
Consider bias from data, algorithmic characteristics, and choice of outcome targets throughout the AI lifecycle
Implement incentive structures that will encourage desired behaviors and processes and promote democratization of health AI
Promote user-centered design of health AI tools and applications to optimize satisfaction, ease of understanding, and appropriate use in health AI applications
Promote local governance that encourages standardization while still supporting customization to the local environmental, cultural, and clinical contexts of use
Create a safety culture, including non-retaliatory reporting for adverse outcomes and includes the training necessary to adapt to an AI-enabled health environment

TABLE 5-3 | Distinct Contributions of Various Stakeholders to the Application of the AI Code of Conduct Framework

Stakeholder	Distinct Contribution
Developers	Developers have vast experience regarding methods and practices, and their active participation in developing standards for the industry will be foundational.
Researchers	Researchers are positioned to provide scientifically sound and independent assessment of both methods and outcomes associated with health AI, including issues of data de-identification, and sharing and the associated implications of societal benefits and burdens, as well as the best practices and standards in workflow integration.
Health Systems and Payors	Local adaptation that facilitates human agency and promotes patient-centered care is the purview of health systems and payors, as is the training and support of the health care workforce in the use of AI in the local health care delivery context. Health systems and payors have an opportunity to create financial incentives that support equitable and effective health AI, using both increases and decreases in reimbursement to support desired best practices around AI use.

continued

TABLE 5-3 | Continued

Stakeholder	Distinct Contribution
Patients and Advocates	Patients, as the recipients of health AI, are uniquely positioned to describe in detail their experience about the impacts of health AI on their lives. Only they can articulate their preferences about critical issues such as access controls over their data or explanations about when and how AI is used in their care. Only patients can share their own personal experiences, both positive and negative, of engagement with developers and the health care system as the use of AI for diagnosis, treatment, and payment advances. And patients are by definition the only source of patient-reported outcomes.
Federal Agencies	The funding and regulatory authority held by the federal agencies has the power to shape the future of health AI. Some examples of how these tools could take form include through support for studies to measure how AI can influence patient health, human agency, goals of care, and human-human interactions in the presence of AI interventions; through recognition of standards for collection and exchange of relevant data and encouraging use of the Trusted Exchange Framework and Common Agreement for making data available for training algorithms, and prioritized research projects; or through the expansion of requirements in AI product labeling based on real-world performance.
Health Care Workforce	As end users of some types of health AI, the health care workforce is situated to identify workflow needs and priorities, and as purchasers or influencers of purchasing decisions, clinicians in particular may have contracting opportunities to require disclosure of AI models' alignment with the Code Principles and Commitments and address liability concerns should model outputs cause harm.
Quality and Patient Safety Experts	Quality and patient safety experts and accrediting organizations play the role of independent auditors, ensuring that processes are designed and implemented and metrics are developed and routinely assessed to ensure the quality of outputs and reduce the risk of harm from health AI tools.
Ethicists and Equity Experts	Ethics and equity experts are uniquely qualified to consider and weigh the novel tensions health AI presents across various stakeholder priorities, always holding the greatest good for the health of the individual and the community as the north star. They are positioned to serve as guides on the path to implementing trustworthy AI.

6
THE COMPONENTS REQUIRED TO ADVANCE RESPONSIBLE AI

Promoting positive change at scale in the use of artificial intelligence (AI) to impact health, health care, and biomedical science is challenging, in part due to the multitude of interacting activities and stakeholders. Advancing successful development, implementation, and sustained use of trustworthy health AI will require careful coordination and consideration of critical elements and how they work together. Identification of activities that are amenable to collaboration, consolidation, and centralization is warranted to ensure both efficiency and quality of outcomes.

Caldwell and colleagues' model for leading change at scale provides a useful organizing framework (Caldwell et al., 2005). The Tight-Loose-Tight leadership model describes how change can be effectively propagated through complex systems while maintaining flexibility, supporting a shared ownership, and ensuring stakeholder engagement at all stages. This model is particularly relevant to advancing responsible AI in the health sector. By design, the Tight-Loose-Tight model balances innovation and control, encourages collaboration and builds trust, and supports an iterative, dynamic, and scalable approach that is needed for emerging technologies subject to multilevel governance. The stages of the model are depicted in Figure 6-1 and are described here in more detail.

- **Tight 1: Define Vision, Values, Goals, and Expectations:** This portion of the model is focused on aligning vision and values through a shared set of principles and commitments (such as the Code Principles and Code Commitments) and on establishing clearly defined goals and expectations as well as desired outcomes and approaches. In the context of health AI, it is driven by key stakeholders, including policy makers, technology firms, health leaders, and various advocacy groups, toward the development and alignment

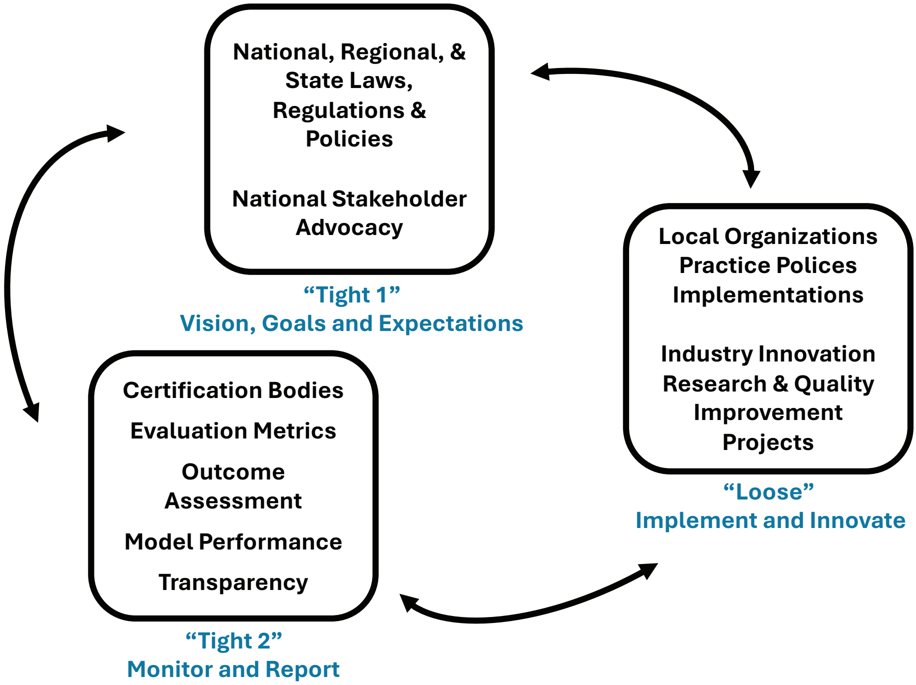

FIGURE 6-1 | Representation of the Tight-Loose-Tight model of leading change at scale adapted for the AI in the health and health care context.
SOURCE: Conceptually adapted from Compton-Phillips, 2019.

of principles, laws, regulations, and standards adopted to reflect desired health outcomes in the context of societal norms and needs.

- **Loose: Implement and Innovate:** This portion of the model addresses the implementation of the goals and vision and the translation of the principles into practice in the context of the laws, regulations, and standards established in the initial phase (*Tight 1*), with the implicit expectation that testing, adaptation, and flexibility will be necessary to address the local context. In the context of health AI, testing of models, local governance, implementation support, research innovation, supports for new technology, and public-private partnerships are included in this phase. It is important that flexibility of thought, practice, and execution be promoted where possible, if fidelity to the overarching requirements described in the *Tight 1* phase is to be maintained and conscious attention to standardization in the *Tight 2* phase occurs as well.

- **Tight 2: Monitor and Report Outcomes:** This third aspect of the Tight-Loose-Tight model is about assessing the degree to which the first two phases have produced the intended outcomes. It represents a more standardized approach to establishing the measurable outcomes and metrics for the process. This phase is focused on assessing whether the health AI processes, tools, and workflows have met the pre-established vision and goals and align or comply with the governing principles, laws, and regulations, including local policies and procedures. It is important that these are reproducible, measurable, transparent, and interpretable, so that all stakeholders involved in the use of health AI understand whether the intended use was achieved.

TIGHT 1: VISION, VALUES, GOALS, AND EXPECTATIONS

Alignment of Health Care AI Principles, Guidelines, and Frameworks Nationally

Internal alignment and coordination on health AI strategies, policies, and relevant regulations across federal agencies are all well underway. As previously noted, in July 2024, the Department of Health and Human Services (HHS) announced a reorganization designed to "streamline and bolster technology, cybersecurity, data, and AI strategy and policy functions" (HHS, 2024b). This reorganization addressed the decentralization of AI policy and operations responsibility, which had been jointly managed by the Office of the National Coordinator for Health Information Technology (ONC), the Assistant Secretary for Administration, and the Administration for Strategic Preparedness and Response. A new role of Assistant Secretary for Technology Policy (ASTP) and ONC was established to consolidate these functions and to provide centralized "oversight of technology, data, and AI policy and strategy" (HHS, 2024b). ASTP ONC established an Office of the Chief Technology Officer, which is responsible for overseeing "cross-agency technology, data, and AI strategy and policy" and includes the Offices of the Chief AI Officer, Chief Data Officer, and Digital Services (HHS, 2024b). This internal alignment is an important step to ensuring consistency regarding vision, values, goals, and expectations for health AI across the federal government.

The AI Code of Conduct (AICC) Code Principles and Code Commitments were drawn from a landscape review encompassing both peer-reviewed literature and federal guidance available at the time of publication (Adams et al., 2024). Similarly, there is an ongoing need for alignment of the myriad sets of principles, guidelines, and frameworks for responsible health AI currently proposed by various national organizations. The Code Principles and Commitments are not intended to be a "one-size-fits-all" vision or goals statement that should be adopted wholesale by other groups, but rather as a reference point to promote national alignment. Wholesale adoption of the Code Principles and Commitments without identifying organizational values and goals would not bring about the same level of commitment, given that people support what they help create (Christiano and Neimand, 2018). During the process of cross-walking with the AICC, in addition to furthering national alignment, organizations and individuals could contribute to the continual improvement of the AICC if they identify gaps or opportunities for revision of the current Code Principles and Code Commitments.

Alignment Among Federal and State Governmental Agencies

An additional need for alignment exists in the context of state and federal legislation and regulation. A patchwork of disparate regulations within the United States can create a costly, complicated compliance environment and present a risk of impeding innovation, in part because small, innovative companies may struggle to track and comply with regulations across all states and territories, and some companies may avoid bringing innovations to states with burdensome regulations. But small innovators and idea incubators contribute significantly to the innovation of new technologies (Kesavan and Dy, 2020). Additionally, fragmented regulations may lead to uneven consumer protections across various jurisdictions as seen with "health settings outside the hospital and clinic" (Aggarwal et al., 2020).

Alignment of federal and state regulations to ensure consistency in AI governance across different jurisdictions could be facilitated by establishing task forces or working groups composed of federal and state legislators and regulators to address AI policy and regulatory issues. These groups could work to ensure ongoing communication, joint decision making, the development of unified or aligned strategies, policies, and regulation, and collaborative learning. They could also develop shared metrics for evaluating the impact and effectiveness of AI regulations, allowing for consistent assessment and improvement (IOM, 2002).

National Standards for Responsible Health AI

Multiple public and private entities such as the U.S. Food and Drug Administration (FDA) and the International Organization for Standardization (ISO) have drafted or produced guidance on AI development and testing (FDA, 2022; ISO, 2023). However, there is currently no governing body responsible for the codification of health AI evaluation standards, resulting in inconsistent evaluations of AI solutions (Saria, 2022). Convergence on shared standards is foundational to consistently optimized outcomes.

A fundamental need is for objective and verified assessment of conformity to standards around processes and outputs, including reports on long-term model performance across various populations and sites. There is an opportunity to learn from the electronic health record (EHR) certification processes established by the ONC. Lessons include the need to (a) test and certify products in settings and on data comparable to real-world applications; (b) address misaligned incentives for certification bodies paid by developers; (c) provide public evidence in support of adherence to standards rather than to rely solely on attestation; (d) recertify periodically; and (e) conduct post-market surveillance to promote maintenance of certification and local governance (Ratwani et al., 2024). There is currently "no publicly available, nationwide approach that enables objective assessment of health AI models and the consequences of their use" (Shah et al., 2024, p. 245). Verifiable, objective, and ongoing standards-driven assessment of health AI models and systems is critical to ensure that the risks of AI in health care and biomedical research are identified and proactively mitigated and to engender trust among all key stakeholders. Such an assessment should be supplemented by the need for an infrastructure for local validation and implementation science framing to ensure effective use of AI.

National Standards for Education About AI

An essential component of establishing shared goals and expectations about health AI is establishing a baseline, a common understanding about what exactly AI is, how it works, what benefits it presents and risks it poses, as well as what accountabilities various parties have in its use. While AI is playing an increasingly important role in health care delivery, AI is typically identified as a gap in medical education (Krive et al., 2023). Establishing admission requirements, core medical education curriculum and/or continuing medical education, and board certification requirements could reduce the knowledge gap in the field (Paranjape et al., 2019). Similarly, guidance around health AI in nursing and allied health training and practice is limited (Glauberman et al., 2023). Collaboration among educational accrediting bodies to establish curriculum standards for medical education and health care professional education in general warrants further consideration.

LOOSE: IMPLEMENT AND INNOVATE

Local Implementation and Governance

Local implementation of health AI applies the vision, values, goals, and expectations set out in the *Tight 1* phase with consideration for the local organizational culture and requirements as well as sub-population patient preferences. It is an exercise in change management and addresses issues including leadership support, alignment with organizational vision and priorities, communication strategies to ensure understanding and ongoing alignment, and establishing goal posts to measure success (Phillips and Klein, 2023).

Broadly, governance is a "systemic, patterned way in which decisions are made and implemented" (European Observatory on Health Systems and Policies, 2016), and local governance typically refers to the applications of policies, procedures, and oversight mechanisms by non-governmental organizations. It includes alignment and compliance with federal, state, and local regulations and contractual requirements, and alignment with national consensus organizations. It provides for ethical oversight and ensures appropriate processes and metrics for testing and monitoring of new and existing systems (Kim et al., 2023). Local governance also defines skill requirements and training expectations for teams using health AI; and establishes authorities and accountabilities for the use of health AI in the local context.

An example of local adaption of AI applications would be for a rural hospital that cares for a large migrant farm-worker community to include workflow elements intended to reflect data collection, patient communication and language preferences, and re-contact needs. The goal would be to achieve pan-stakeholder agreement on health outcomes. Another example is an urban setting in which transportation, access, and health care delivery trust are primary care barriers, and an emphasis on adapting AI applications to focus on patient needs and preferences in these settings, aimed at promoting human health is warranted.

There has been considerable focus on the development of frameworks and maturity models to measure the capability of health care delivery systems to deploy, use, and monitor AI solutions frameworks, including national coordinating groups such as the Health AI Partnership (HAIP, n.d.) and the Coalition for Health AI (CHAI, n.d.), as well as commercial entities such as MITRE (2024), and large clinical systems such as Duke AI Health (n.d.). However, there has been limited reflection on the whole of health AI governance, and there are currently no national standards against which organizations can measure their performance on all aspects of local health AI governance. While some of these standards could be addressed by the Centers for Medicare & Medicare Services (CMS) safety requirements for health systems (Fleisher and Economou-Zavlanos, 2024), there remain safety challenges health AI that are likely to require AI-specific safety controls. As these controls mature, they could be incorporated into expectations set by accrediting bodies such as The Joint Commission or the National Committee for Quality Assurance. Within the context of health AI, components warranting consideration for maturity models include presence and adherence to organizational policies and procedures, a program for monitoring outcomes, clear accountability structures, and training and communication plans.

Technical Implementation Support for Under-Resourced Health Care Organizations

When considering the establishment of sound health AI governance and the adoption of health AI, local context is important. For under-resourced organizations such as those that are small, that function in rural settings, or that serve populations at risk, the ability to attend to the complexities of health AI is limited when those organizations are compared to larger and better-resourced entities. This could lead to such systems not adopting health AI or being unable to properly control the management of health AI adoption and use; either case could result in an expansion of disparities and/or the digital divide. A similar risk was identified during the adoption and use of standardized health information technology (IT) such as EHRs by health care organizations. As noted earlier in this chapter; to address this risk, Congress used the Health Information Technology for Economic and Clinical Health (HITECH) Act to direct HHS to establish the Health IT Regional Extension Center (REC) Program (ASTP ONC, n.d.). The $720 billion REC program was aimed primarily at helping providers who were treating underserved populations, promoting dissemination of information and assistance on best practices for health IT adoption and ongoing use. The types of local entities receiving REC funds to train providers included health IT research and consulting organizations, universities, quality improvement organizations, and health center networks (Lynch et al., 2014).

Most providers who received assistance from RECs had adopted EHRs, and nearly half had met the criteria for "meaningful use" for obtaining the HITECH federal incentive dollars (Farrar et al., 2015). A similar program with funding to support development and dissemination of best practices for health AI could help entities (under-resourced entities in particular) to develop and implement strategies for deploying responsible AI.

Promotion of Innovation

Novel AI tools offer a transformational opportunity to improve human health and to create meaningful value for all key stakeholders. Flexibility in the *Loose* phase is required to foster innovation supported by a culture of creativity, collaboration, and measured risk-taking, as well as rigorous assessment, all conducted within the parameters set forth in the *Tight 1* phase. Advances can be derived from efforts and investment by public agencies, private organizations, or public-private partnerships.

In the context of public efforts and investment, federal agencies engaged in the support and conduct of research can provide guidance on the needs, gaps, and priority areas in health AI research and fund special initiatives and consortia for critically needed research that may not be feasible in a small setting. For example, HHS's Advanced Research Projects Agency for Health is seeking to develop calls for research proposals and funding to develop democratized real-world health care data sources from EHRs, claims data, social determinants of health, and environmental data, among others. They are also seeking to provide mechanisms to use these data for AI training through federated mechanisms that facilitate wider and more equitable AI tool development (ARPA-H, 2024). Another example of this is the National Institutes of Health's (NIH's) Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD, n.d.) whose focus is to advance health equity and develop capacity for AI experts among underserved populations. At a more foundational level, the National Science Foundation has been funding a wide range of AI research projects through many initiatives, including the National Artificial Intelligence Research Institutes (NSF, n.d.). Some of the innovations from these initiatives have diffused into the health space and are driving some of the innovation in this sector. Ongoing funding for robust scholarly study is essential to build the needed evidence base to advance AI for human health.

Big tech has poured hundreds of billions of dollars into AI research and development (Ahmed et al., 2023b). However, as noted earlier, while small innovators and idea incubators generate important solutions and new technologies (Kesavan and Dy, 2020), they are often unable to compete at scale, and as a result, their advances may not be realized or may be delayed in the market. It may be beneficial to consider developing governmental programs that support entrepreneurs, start-ups, and small businesses, particularly minority-owned, as an alternative to those that require such commitments as an equity stake in the company or a share of future revenues.

Public-private partnerships offer another means of promoting innovation in health AI. Of particular interest is the role of government agencies partnering with private organizations (such as academic institutions and/or technology companies) "to facilitate the translation of health data into actionable insights to streamline operations, improve care coordination, and enable greater insights" (Arnaout et al., 2023). Public–private partnerships can yield innovation that would otherwise be impossible, impractical, excessively time consumptive, or cost-prohibitive for any individual academic group, private organization, or government agency, and as such, provide an important path forward to advancing health through AI.

Regulatory Innovation

Rapidly evolving technologies can present challenges for regulatory bodies, both in assessing whether and how existing laws, rules, and frameworks apply to the new technology and in responding thoughtfully and rapidly to identified gaps. In this *Loose* phase of change at scale, innovation in policy making through regulatory sandboxes presents an opportunity to mitigate the risks associated with unintended consequences of maintaining current policy by granting individual exceptions to existing regulations, allowing policy makers to monitor the differences in outcomes that current rules yield relative to the exception (Leckenby et al., 2021). The results of these exceptions then inform legislative or regulatory changes based on "more robust regulatory knowledge" (Buocz et al., 2023). Although they have been used in regulation of financial technology, regulatory sandboxes also present an important opportunity in areas such as health AI to inform policy making, to reduce time to market for new products, and to promote greater safety for consumers (OECD, 2023).

Another innovative regulatory approach has been proposed by Blumenthal and Patel (2024) who posit that the novel challenges of generative AI (GAI) require special consideration for governance. Given a general-purpose technology that is subject to model drift, and which may produce unreliable results, they suggest "one possible direction would be to treat these GAI-based clinical applications not as devices but as a new type of clinical intelligence" (Blumental and Patel, 2024). This novel approach would require managing these GAI–based tools much the way a clinician is treated, through proper training, testing, supervision, ongoing retraining, periodic reporting on quality to regulatory authorities, and so forth.

Financing AI in Health Care Delivery Systems

While the integration of AI into health care delivery systems holds much promise, it is also fraught with financial challenges that raise significant concerns. Testing various means of paying for health AI is consistent with the local application of new technology in the *Loose* phase. The high initial costs of acquiring AI tools, the ongoing expenses related to maintenance and updates, and the need for specialized personnel to manage these systems represent substantial financial burdens. Moreover, the disparities in financial resources among various stakeholders can exacerbate existing inequalities, leading to unequal access to cutting-edge AI innovations. Foregoing some type of financial support for AI may not be desirable. Like telemedicine, without adequate reimbursement for AI in its early stages, longer-term benefits of AI for health outcomes and operational efficiency may not be realized (Parikh and Helmchen, 2022). The pressure to demonstrate return on investment is significant; AI has been used by providers and payors to improve their financial position. Providers have used AI to maximize revenue by optimizing coding, improving billing accuracy and completeness, and streamlining the prior authorization processes (Zhu et al., 2024). Simultaneously, much attention has been given to using AI to detect fraud in medical billing (du Preez et al., 2024). Payors have used AI to automate claims adjudication, resulting in more and more-rapid denials, leading to iterative and escalating responses between providers and payers (Williams, 2024). The application of the Code Commitments to Advance Humanity and Ensure Equity by all parties will be necessary to ethically address competing priorities during this *Loose* phase.

Some AI tools may offer benefits that are not immediately reflected in a health system's bottom line, making the investments more challenging to justify from a purely financial perspective. For example, AI tools can analyze a patient chart and identify pertinent information for a clinician, potentially significantly reducing provider time for record review, and thereby potentially improving clinician well-being (Parikh and Helmchen, 2022). These financial hurdles necessitate an exploration of potential funding models, cost-benefit analyses, and strategies for ensuring equitable access to these tools by health systems. Ensuring that the advantages of AI in health care delivery are accessible to all, regardless of financial standing, is crucial for fostering innovation and assuring equitable distribution of the benefits.

Costs Associated with Implementing AI in Care Delivery

Estimates of the costs of investment by health sector stakeholders in implementation of AI, for either in-house solutions or purchased systems, are opaque and dependent on the specific applications deployed. For example, the total cost of deployment of a chatbot might be several thousand dollars per year, while the total cost, including integration of a complex predictive model into an existing system could cost hundreds of thousands of dollars over time (Sanyal, 2021). Many cost estimates consider only the up-front costs. A comprehensive understanding of the total cost of AI systems is important and includes not only up-front costs, but also integration with existing systems, ongoing data management and security, system monitoring and maintenance, staff acquisition and training, and regulatory compliance.

Financial Benefits Associated with Implementing AI in Care Delivery

A major driver of the use of AI in the sector is clearly improved health. Beyond improved clinical outcomes, economists have predicted significant financial benefits from the use of health AI, including one estimate of a 5–10% reduction in annual health care spending (Sahini et al., 2023). Anticipated efficiencies yielding cost reduction opportunities abound. Using AI, health systems may improve efficiency in patient–provider matching, scheduling, referrals, and reduction of missed appointments. AI-powered clinical decision support systems can speed health care professionals' workflows by providing real-time recommendations based on evidence-based guidelines (Sutton et al., 2020).

Beyond self-funded AI implementation, several innovative approaches are being tested or could be tested in this *Loose* phase to support equitable access to the benefits of health AI. CMS is currently approaching AI reimbursement in three ways: (1) simply including AI as part of existing payments for services, (2) providing an additional transitional payment, or (3) creating a new procedure and paying for it independently of other services (Zink et al., 2024). CMS could incorporate AI costs into existing bundled payments for services while modifying the price of the bundled reimbursement based on quality of the AI and costs for its development, balancing innovation and affordability for clinicians (Zink et al., 2024). AI device manufacturers could consider offering volume-based pricing and accept downside risk if patient outcomes are not realized; or, gain-sharing models could be established between device manufacturers and clinical systems (Parikh and Helmchen, 2022). Expansion of effective approaches and/or the development of novel reimbursement strategies should be pursued, ensuring that they balance innovation and uptake with responsible, value-based use in practice.

TIGHT 2: MONITOR AND REPORT OUTCOMES

Evaluation Metrics

During this phase of scaling health AI, the objective is to assess the performance of health AI in the local context, considering adherence to stated processes and requirements, as well as performance relative to intended outcome goals identified in the *Tight 1* phase. As stated earlier, reproducible, measurable, transparent, and standardized, interpretable metrics are needed.

Health AI applications cover a broad range of technologies, from traditional predictive models to more evolutionary GAI. Each such domain requires tailored evaluation methods, yet the absence of standardized metrics leads to fragmented and difficult-to-compare assessments. This diversity complicates direct comparisons and undermines the consistency of performance claims. While it may not be possible to specify the exact metric applicable to a specific use case and no single standardized metric can be used to assess performance of all health AI applications (Hicks et al., 2022), there are a number of frameworks by which metrics can be deemed sufficient and the processes used to establish and evaluate health AI, and the downstream actions, surveillance, and eventual decommissioning standardized.

FDA (2021) and the European Medicines Agency (EMA, 2021) provide some relevant guidelines for AI evaluation; however, these are not universally applicable or consistently used across different types of AI systems. And, in 2023, an international ISO standard was established for the management of AI technologies in practice, and intended to address challenges in ethics, transparency, and continuous learning; it specifically includes sections on performance evaluation and improvement (ISO, 2023). However, this standard is high level and is intended to be a starting framework to grow as AI maturity grows.

The Joint Commission's Responsible Use of Health Data certification program assesses controls for secondary uses of health care data (including AI training or local validation) in areas of de-identification, data controls, limitations of use, and transparency to patients, as well as the process an organization has put into place for validation of any data-based algorithm.

The World Health Organization has published guidance on standards for evaluation of health AI for medical devices (WHO, 2021); this guidance is relevant to health AI more generally and local implementation specifically. It includes consideration of *discriminating measures* capable of identifying patients who will, versus those who will not, have an adverse event, as well as *calibrating measures* capable of assessing the accuracy of risk prediction (Alba et al., 2017). The WHO model also calls for evaluation of AI performance on sub-populations as appropriate as well as for comparison outcomes relative to care delivered without AI. Furthermore, the model calls for careful consideration of issues such as sample size, inclusion criteria, and follow-up periods, and at a minimum recommends measurement of clinical processes and outcomes, potential harms, user and recipient experience, and comparisons to gold standard.

Robust local evaluation of implemented AI systems is essential to ensuring that health goals are achieved, and trust is built and maintained. Collaborative development, broad stakeholder engagement, and standards alignment between and among organizations and agencies seeking to govern or accredit local health AI users warrants careful consideration. Specific issues that must be addressed include those that follow.

Reliability and Validity

Shared, standardized performance metrics are essential for ensuring the reliability and validity of health AI systems. These metrics ideally encompass various dimensions of performance, including accuracy, generalizability, interpretability, and fairness (Kiseleva et al., 2022; McCradden et al., 2023). By standardizing such measures, stakeholders can more accurately gauge and understand the effectiveness of AI applications, leading to more informed and empiric decision making when selecting, designing, optimizing, and monitoring AI platforms.

Accuracy and Generalizability

Accuracy, a critical performance metric, determines the correctness of AI predictions or classifications. Generalizability, on the other hand, assesses the system's ability to maintain performance under varying conditions (Daneshjou et al., 2021; Fehr et al., 2024). Standardizing these metrics would allow for consistent benchmarking, enabling stakeholders to identify the most reliable systems.

Interpretability and Fairness

Interpretability refers to the ease with which stakeholders can understand and trust AI decisions. Fairness ensures that AI applications do not perpetuate or exacerbate biases in underlying data or as encoded in algorithms (Fehr et al., 2024; Vollmer et al., 2020). Again, establishing shared metrics for these dimensions of health AI is crucial for enhancing trust and ensuring ethical AI deployment in various clinical settings.

AI Model Performance Transparency

As is the case for AI evaluation metrics, transparency standards are essential to demonstrating efficacy and building trust, and such standards currently do not exist, from the perspective of either researchers or end users or the recipients of AI. With an agreed-upon set of standard performance metrics, dissemination of results will be important for various audiences. An initial set of requirements for algorithm transparency came from ASTP ONC in the form of transparency requirements for algorithms incorporated into certified health IT (ONC, 2024a).

Additionally, reporting frameworks for AI-focused clinical studies that have been adopted by major peer-reviewed publications in biomedical and life sciences. Such reporting standards emphasize transparency, reproducibility, and clarity in the documentation of AI methodologies and results. Key frameworks, such as Consolidated Standards of Reporting Trials for AI and Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence (SPIRIT-AI), provide guidelines for reporting clinical trials involving AI, focusing on detailed descriptions of the AI intervention, data pre-processing, model training, and validation processes (Ibrahim et al., 2021). These standards also stress the importance of disclosing performance metrics, such as accuracy, sensitivity, specificity, and any potential biases. Furthermore, they require authors to describe the clinical context, intended use, and limitations of the AI system, ensuring that peer reviewers and readers can critically assess the validity and applicability of the research findings.

In the context of a clinician assessing an individual algorithm, model cards or labels may provide a template for ensuring transparency (Sendak et al., 2020). The proposed model labels aim to provide frontline clinicians with an easy-to-digest, one-page document with the necessary information to help end users to know when and how to apply AI model output in decision making. In addition to statistical performance of the AI, model, facts include essential information such as an overview of the model, outcomes and outputs, data sources, uses and directions, and warnings.

From the perspective of patients and their advocates, transparency is essential to ensure clinical benefits and mitigate risk of harms (The Light Collective, 2024). The group is calling for transparency about why and how their data are being used in AI models; when AI is guiding their care; and what the level of evidence is to support the use of AI in their care.

Capacity for Shared Learning Between and Among All Stakeholders

A feedback loop from the *Tight 2* local context (bottom up) to the *Tight 1 broader governance context* (top down) stages of change at scale is essential to ensure ongoing learning and adaption across the whole of the health ecosystem. There are a variety of capacity-building approaches to consider. Interdisciplinary coalitions, collaboratives, and networks may hold joint meetings, workshops, or conferences to exchange insights, recognize shared obstacles, and create joint solutions.

Collaborative research projects involving health care providers, patients, AI developers, and academic institutions offer an opportunity to explore AI applications and their implications in clinical settings. Funding specifically in support of multidisciplinary collaboration could promote diverse perspectives in AI health care research.

Shared learning could also be advanced through online learning platforms with e-learning modules tailored for different stakeholder groups, including clinicians, developers, and patients and through communities of practice where stakeholders could share experiences, best practices, and resources related to AI in health care.

Additional opportunities include pilot projects with planned and structured debriefing sessions to capture learnings and insights and simulation-based training to allow stakeholders to experience AI-driven health care scenarios. Knowledge brokers could bridge the gap between AI developers and health care practitioners, translating technical AI knowledge into practical insights for clinical use. And finally, it is essential that researchers and AI practitioners regularly publish their findings from AI health care initiatives in accessible formats, such as white papers, case studies, and reports, to inform and educate stakeholders.

Key components for successful implementation of safe, effective, trustworthy health AI can be considered from a Tight-Loose-Tight framework, summarized in Table 6-1, where broad, shared agenda setting is followed by local implementation and innovation, and then monitoring and reporting of outcomes for shared learning. In the *Tight 1* phase alignment on vision, values, goals, and expectations can be advanced through state, federal, international, and public–private collaboration on shared standards and assurance. Innovation and implementation in the *Loose* phase can be supported through local governance, public and private investment, and regulatory experimentation. Finally, standardized evaluation metrics, transparency, and ongoing feedback and shared learning are needed to ensure success in the *Tight 2* phase to promote change at scale. Critical collaboration and work remain for all stakeholders across all phases of change to ensure that the benefits of health AI are realized.

TABLE 6-1 | Summary Key Components to Advance Health AI via the Tight-Loose-Tight Framework

Key Components to Advance Health AI			
	Tight 1	*Loose*	*Tight 2*
Goal	Align vision, goals, and expectations	Innovation and implementation	Promote change at scale
Actions	State, federal, international, and public–private collaboration on shared standards and assurance	• Local governance • Public and private investment • Regulatory experimentation	• Standardized evaluation metrics • Transparency • Ongoing feedback and shared learning
Stakeholders	• Policy makers • Technology firms • Health leaders • Advocacy groups	• Local organizations • Industry	• Guideline and standard makers • Regulators
Priorities	• Alignment of health care AI principles, guidelines, and frameworks • Alignment among federal and state governmental agencies • National standards for responsible health AI • National standards for education about AI	• Local implementation and governance • Technical implementation support for under-resourced health care organizations • Promotion of Innovation • Regulatory sandboxes • Financing AI in health care delivery systems	• Standardized evaluation Metrics ○ *Reliability and Validity* ○ *Accuracy and Generalizability* ○ *Interpretability and Fairness* • AI model transparency • Capacity for shared learning between and among all stakeholders

7
ADVANCING THE AI CODE OF CONDUCT FRAMEWORK: THE "VITAL FEW" PRIORITY ACTIONS

To realize the benefits and avoid the risks associated with health artificial intelligence (AI), it is imperative strategically to prioritize key actions that are most likely to ensure that incentives and supports are intentionally designed and properly executed, and that progress is both effective and responsive to a changing environment. This will require significant effort from and coordination between all stakeholder groups. In addition, beyond the targeted efforts outlined below, much work will be required by multiple parties. The actions listed are not intended to be comprehensive but instead constitute the "vital few" highest priorities to advance the Code Principles and Commitments, building the capability to rapidly respond to new tools, technologies, opportunities, and concerns.

In identifying priority actions for the translation of the Code Commitments to real-world application, information synthesis, gaps, and opportunities were identified by all contributors to this work. Two additional constructs were considered; the first was the identification of priority actions that are foundational to, supportive of, or capable of catalyzing additional needed action, thereby likely to create a cascading effect and potentially speeding the national collective effort to promote safe, effective, and trustworthy health AI. The second was the application, as appropriate, of behavioral economics, which posits a set of strategies including incentives, defaults, and framing to make preferred decisions and actions easier and more rewarding or less costly (Siegel et al., 2021). Below, for each commitment, priority actions are outlined for consideration in the application of the Code Commitments in real-world settings.

COMMITMENT 1: ADVANCE HUMANITY

To ensure that human health, agency, and connection remain the primary focus of health AI, it is essential to identify, transparently characterize, and promote the societal and cultural goals of the recipients of the use of health AI as an accountable mechanism to protect and advance human health and connection.

Development of standards and other governance structures to assess alignment by developers and users of health AI with societal and cultural goals for health AI:

- Requirements to understand end user and recipient needs and preferences.
- Methods at all governance levels to assess the degree to which innovation maintains a foundation of beneficence and the avoidance of harm.
- Patients and patient advocacy groups as central, engaged stakeholders with agency to ensure that this group has the capacity to influence decisions and outcomes in all aspects of health AI.

Incentives and structures for independent evaluation, certification to the AI Code Commitments, and public and transparent reporting on certification status:

- Proactive information disclosure to intended audiences regarding the development and use of AI, actively involving them in the planning process, educating them, and aligning with their needs and preferences.

COMMITMENT 2: ENSURE EQUITY

To ensure equitable distribution of benefits and risks of health AI, it will be critical to place equity and fairness at the center of all health AI development and use and ensure its prioritization throughout the AI lifecycle.

Standardized metrics to assess and report bias in data, AI output, and AI use, in the interest of equitable distribution of benefit and risk:

- Diverse, representative data sources for AI training and local implementation.
- Proactive and reactive bias mitigation.
- Metrics assessing fairness and action based on the results.

Incentives and support for low-resourced organizations and communities to ensure equitable access to the benefits of AI:

- Incentives and supports akin to those provided to drive electronic health record adoption.
- Alignment of incentives and payment models to promote equity and democratization of health AI.

COMMITMENT 3: ENGAGE IMPACTED INDIVIDUALS

To ensure that key stakeholders are viewed as partners with agency in every stage of the lifecycle, it is important to identify, engage, and most importantly, *integrate* all relevant stakeholder input in conceptualization, design, development, implementation, and surveillance throughout the health AI lifecycle.

Participation by all key stakeholders across the health AI lifecycle:

- Engagement with intended users of health AI to ensure preferences and workflow integration where possible.
- Engagement of ethics and equity experts in all AI development.
- Expanded stakeholder inclusion in developing research program funding and goals, which federal agencies and other funding entities could consider.

Local governance bodies that include all stakeholders in the AI lifecycle:

- Local governance frameworks and maturity models.
- Transparent processes for decision making, conflict management, and prioritization to review, understand, and redress stakeholder conflicts, using the Code Principles and Code Commitments as guidance.

Common understanding/education of all impacted parties:

- Patient and end-user education programs designed in collaboration with them, and in a culturally appropriate fashion to support awareness of risks and benefits of health AI and support their personal decision making about its use in their process of care.
- Appropriate communication channels to reach patients to promote a shared understanding of what AI is, what the risks and benefits of AI use are, how it is used for their care, and how they can access available information on AI transparency (akin to ClinicalTrials.gov).

COMMITMENT 4: IMPROVE WORKFORCE WELL-BEING

Consistent with the priorities laid out in the National Academy of Medicine Plan for Health Workforce Well-Being (NAM, 2022c), it is imperative to create a shared sense of purpose and potential for the health care workforce. Top priorities include workforce education and investment in research.

Positive work and learning environments and culture (NAM, 2022c):

- Curriculum standards for trustworthy AI competencies across the spectrum of stakeholders including developers, researchers, and the health care workforce.
- Well-defined communication and educational programs to support health care workforce AI competence and comfort in use.
- Programs and evaluation mechanisms for workforce reskilling to facilitate retention, adaptation of roles, and implementation of change management processes to ease the introduction of disruptive technologies.
- Training and expansion of the health care workforce that is AI aware and AI competent.

Measurement, assessment, strategies, and research (NAM, 2022c):

- Advancement of science and understanding of the interaction of AI technologies and health care delivery workflow.
- Change management tools and techniques to develop capacity for workflow redesign.

COMMITMENT 5: MONITOR PERFORMANCE

Effective monitoring and sharing of AI's performance and impact on health and safety will require stakeholders to integrate and align risk management and quality measurement and assurance frameworks for the health AI lifecycle. Careful consideration is needed to assess technical rigor, use case utility, and trustworthiness (equity, fairness) in the conduct of performance monitoring.

Standardized quality and safety metrics to assess the impact of the use of health AI on health outcomes:

- Processes that can document and report on technical and clinically meaningful performance metrics that are understandable to the public, patients, and their caregivers.

- Requirements and frameworks regarding health AI performance monitoring that balance safety, health utility, and flexibility.

COMMITMENT 6: INNOVATE AND LEARN

Innovation and discovery are needed to drive continuous improvements to health, and shared learning and ongoing systems feedback are the foundation of the Learning Health System.

A well-supported national health AI research agenda:

- Federal research, development, and implementation efforts to support an ecosystem of safe and effective health AI.
- Collaborative research projects involving health care providers, patients, AI developers, and academic institutions that explore AI applications and their implications in clinical settings.
- Efforts to assess not only development and implementation but also the assessment of decommissioning activities.

Participation in shared learning across all stakeholders:

- Meetings, workshops, or conferences to exchange insights, recognize shared obstacles, and create joint solutions.

Innovation as a core investment:

- Entrepreneurs, start-ups, and small businesses that deliver novel solutions to problems in health, health care, and biomedical science.
- Investment in research and advancement in science among federal agencies for critical needs in health AI where commercial stakeholders may have less focus.
- Ongoing support for public-private partnerships as a means of promoting innovation in health AI.

While there is clearly much work to be done across stakeholders to advance responsible health AI, prioritizing actions that impact the highest points of leverage and that are in some cases already in motion will allow us to reap the benefits and avoid the pitfalls of health AI most expeditiously, effectively, safely, and effectively. Applying the concepts of behavioral economics, it will be important to make it easy and rewarding to abide by the shared vision, values, goals, and expectations described in the nationally aligned AI Code Principles and Code Commitments. Established standards, incentives, and transparent performance metrics will be key. Table 7-1 summarizes priority actions to operationalize the AICC Code Commitments.

TABLE 7-1 | Summary Priority Actions to Operationalize the AICC Code Commitments

Key Actions to Operationalize the AICC Code Commitments		
Commitment	Action	Involved Parties
Advance Humanity	• Support the development of governance standards for AI alignment with societal goals. • Create incentives and structures for independent evaluation, certification to the Code Commitments, and public and transparent reporting on certification status.	• Developers; health systems and payors; researchers; ethicists; professional associations, state, federal, and international governments; patients, families, and communities • Federal agencies including ASTP ONC, FDA, NIH
Ensure Equity	• Establish a standard set of metrics to be used proactively and reactively to assess bias in data, AI output, and AI use, and report publicly and transparently. • Provide incentives and supports to low-resourced organizations and communities to ensure equitable access to the benefits of AI.	• Researchers and federal agencies • Federal agencies including ASTP ONC, FDA, NIH, HRSA
Engage Impacted Individuals	• Promote and (incentivize as appropriate) participation by all key stakeholders across the health AI lifecycle. • Establish local governance bodies which includes all stakeholders in the AI lifecycle. • Ensure common understanding/ education of all impacted parties.	• Federal agencies including ASTP ONC, FDA, NIH, HRSA • Developers; health systems and payors; researchers; ethicists; professional associations, state, federal, and international governments; patients, families, and communities • Developers; health systems and payors; researchers; ethicists; professional associations, state, federal, and international governments; patients, families, and communities

continued

TABLE 7-1 | Continued

Key Actions to Operationalize the AICC Code Commitments		
Commitment	Action	Involved Parties
Improve Workforce Well-Being	• Create and sustain positive work and learning environments and culture (NAM, 2022c). • Invest in measurement, assessment, strategies, and research (NAM, 2022c). • Develop workforce AI competency through reskilling and training programs. • Promote well-being by addressing disruptive technologies with change management strategies.	• Developers, health systems and payors, researchers • Developers, health systems and payors, researchers, and federal agencies (e.g., NIH) • Researchers, educational institutions, federal agencies (e.g., Department of Education), professional societies • Health systems and payors
Monitor Performance	• Establish a set of standardized quality and safety metrics to be used to assess the impact of the use of health AI on health outcomes. • Align frameworks to ensure safety, equity, and quality in AI performance.	• Federal agencies, researchers, accreditation bodies, patient safety organizations • Federal agencies, researchers, accreditation bodies, patient safety organizations
Innovate and Learn	• Establish and fund a national health AI research agenda. • Incentivize participation in shared learning across all stakeholders. • Invest in innovation.	• Federal agencies (e.g., NIH) and researchers • Federal agencies (e.g., ASTP ONC)

NOTE: ASTP ONC = Assistant Secretary for Technology Policy, Office of the National Coordinator for Health Information Technology; FDA = U.S. Food and Drug Administration; HRSA = Health Resources and Services Administration; NIH = National Institutes of Health.

8
CONCLUSION

In the past several years, along with the rapid advances in artificial intelligence (AI) methods and technologies, there has been a concomitant spread of novel tools to support health, health care, and biomedical science. These AI tools offer the potential for transformational change in the prevention, detection, diagnosis, and treatment of disease. However, the fundamental differences in these technologies, relative to traditional health information technology, present new risks that require new governance frameworks and new approaches to learning. To that end, the National Academy of Medicine (NAM), under the direction of the AI Code of Conduct (AICC) steering committee, sought to align the health field and to catalyze collective action to ensure that the transformative potential of health AI is realized. The AICC's Code Principles are mapped to the NAM's Learning Health System (LHS) Shared Commitments and a set of simple rules, the Code Commitments, have been developed to aid in the employment of the principles in daily practice. With these aligned vision and goals, as powerful AI technologies and tools advance, the realization of the long-envisioned LHS, is more imminently in reach.

However, AI can be likened to another powerful tool, a hammer, which embodies the duality of utility and intent. The hammer can be used to construct a school, a hospital, a cathedral—each the fulfillment of the collective aspirations of a community. Yet, the same implement can be employed to force incongruity, as in driving a square peg into a round hole, or to engage in a nefarious act, as in breaking a window to commit larceny. So it is with AI, that the value of the technologies and tools are defined by the purpose for which they are directed and governed. Here, the AI Code Principles and Code Commitments offer guideposts.

While health AI is on the cusp of supporting major breakthroughs in biomedical research, medical diagnosis and treatment, maintenance of health, and addressing health workforce challenges, much work remains to ensure that it meets its potential. To meet the moment, all key stakeholders must play a role in ensuring that health AI advances humanity and avoids the risks associated with incongruent or malicious use of the tools and technologies. Ongoing collaboration across sectors is required to articulate the collective aspiration of the community and to advance AI governance. Components of that shared responsibility include:

- Promoting ongoing local, state, national, and international alignment of health AI governance frameworks, including the AICC framework;
- Applying and adapting the AICC Code Principles and Code Commitments to the local context;
- Translating the AICC framework into sector-specific implementation guides;
- Identifying and testing key use cases for application of the AICC framework;
- Establishing local, state, and national policies to support identified priorities; and
- Preparing the workforce to succeed in an AI-enabled environment.

While many technological advances have had potential to improve human health and well-being, they have failed to fully deliver on their promise. Likewise, the realization of a national LHS at scale has remained elusive, in part because contemporary U.S. health care is marked by misaligned incentives, fragmentation, discontinuity, and inundation. Intelligent applications of AI and collaborative governance can serve to bridge many of these gaps and bolster the sociotechnical infrastructure, moving away from fragmented and inchoate data to a system of cohesive and actionable knowledge. The AI Code Principles and Code Commitments can serve as guideposts for decision making about the application of AI tools, and more broadly, prioritizing these fundamental commitments can serve as a framework for increasing trustworthiness, enhancing organizational culture and synergy, accelerating discovery, and promoting effective care. Taken together, advances in health AI delivered in the context of shared principles and commitments can both fulfill the needs expressed in the earliest articulations of the LHS—continuous learning through a dynamic approach to evidence generation and improvement—and, most importantly, realize the shared goal of the health system writ large, to reduce suffering and create health for individuals, communities, and the population as a whole.

REFERENCES

Abdelhady, A. M., and C. R. Davis. 2023. Plastic surgery and artificial intelligence: How ChatGPT improved operation note accuracy, time, and education. *Mayo Clinic Proceedings: Digital Health* 1(3):299–308. https://doi.org/10.1016/j.mcpdig.2023.06.002.

Access Now. 2020. *Access Now Resigns from the Partnership on AI*. https://www.accessnow.org/press-release/access-now-resignation-partnership-on-ai (accessed April 4, 2025).

Adam, G. A., C. K. Chang, B. Haibe-Kains, and A. Goldenberg. 2020. Hidden risks of machine learning applied to healthcare: Unintended feedback loops between models and future data causing model degradation. *Proceedings of Machine Learning Research* 126:710–731.

Adams, L., E. Fontaine, S. Lin, T. Crowell, V. C. H. Chung, and A. A. Gonzalez. 2024. Artificial intelligence in health, health care, and biomedical science: An AI code of conduct principles and commitments discussion draft. *NAM Perspectives*. Commentary, National Academy of Medicine, Washington, DC. https://doi.org/10.31478/202403a.

Adler-Milstein, J., N. Aggarwal, M. Ahmed, J. Castner, B. Evans, A. Gonzalez, C. A. James, S. Lin, K. Mandl, M. Matheny, M. Sendak, C. Shachar, and A. Williams. 2022. Meeting the moment: Addressing barriers and facilitating clinical adoption of artificial intelligence in medical diagnosis. *NAM Perspectives*. Discussion Paper, National Academy of Medicine, Washington, DC. https://doi.org/10.31478/202209c.

Adus, S., J. Macklin, and A. Pinto. 2023. Exploring patient perspectives on how they can and should be engaged in the development of artificial intelligence (AI) applications in health care. *BMC Health Services Research* 23:1163. https://doi.org/10.1186%2Fs12913-023-10098-2.

Advanced Research Projects Agency for Health (ARPA-H). 2024. *RFI: Real World Data Sources for Health and Health Care Research*. https://arpa-h.gov/news-and-events/rfi-real-world-data-sources-health (accessed April 4, 2025).

Agency for Healthcare Research and Quality (AHRQ). 2022. *Improving Healthcare Through AHRQ's Digital Healthcare Research Program*. https://digital.ahrq.gov/sites/default/files/docs/page/ahrq-dhr-2022-year-in-review.pdf (accessed July 3, 2024).

Aggarwal, N., M. Ahmed, S. Basu, J. J. Curtin, B. J. Evans, M. E. Matheny, S. Nundy, M. P. Sendak, C. Shachar, R. U. Shah, and S. Thadaney-Israni. 2020. Advancing artificial intelligence in health settings outside the hospital and clinic. *NAM Perspectives*. National Academy of Medicine, Washington, DC. https://doi.org/10.31478/202011f.

Ahmed, M. I., B. Spooner, J. Isherwood, M. Lane, E. Orrock, and A. Dennison. 2023a. A systematic review of the barriers to the implementation of artificial intelligence in healthcare. *Cureus* 15(10):e46454. https://doi.org/10.7759%2Fcureus.46454.

Ahmed, N., M. Wahed, and N. C. Thompson. 2023b. The growing influence of industry in AI research. *Science* 379(6635):884–886. https://doi.org/10.1126/science.ade2420.

Aikens, R. C., J. H. Chen, M. Baiocchi, and J. F. Simard. 2024. Feedback loop failure modes in medical diagnosis: How biases can emerge and be reinforced. *Medical Decision Making* 44(5):481–496. https://doi.org/10.1177/0272989X241248612.

Ajzen, I. 1985. From intentions to actions: A theory of planned behavior. In *Action Control: From Cognition to Behavior*, edited by J. Kuhl and J. Beckmann. Springer, pp. 11–39.

Alba, A. C., T. Agoritsas, M. Walsh, S. Hanna, A. Iorio, P. J. Devereaux, T. McGinn, and G. Guyatt. 2017. Discrimination and calibration of clinical prediction models: Users' guides to the medical literature. *JAMA* 318(14):1377–1384. https://doi.org/10.1001/jama.2017.12126.

Albahri, A. S., A. M. Duhaim, M. A. Fadhel, A. Alnoor, N. S. Baqer, L. Alzubaidi, O. Albahri, A. Alamoodi, J. Bai, A. Salhi, J. I. Santamaría, C. Ouyang, A. Gupta, Y. Gu, and M. Deveci. 2023. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion* 96:156–191. https://doi.org/10.1016/j.inffus.2023.03.008.

Alizadeh, A., M. Eisen, R. Davis, C. Ma, I. Lossos, A. Rosenwald, J. Boldrick, H. Sabet, T. Tran, X. Yu, J. Powell, L. Yang, G. Marti, T. Moore, J. Hudson, Jr., L. Lu, D. Lewis, R. Tibshirani, G. Sherlock, W. Chan, T. Greiner, D. Weisenburger, J. Armitage, R. Warnke, R. Levy, W. Wilson, M. Grever, J. Byrd, D. Botstein, P. Brown, and L. Staudt. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403(6769):503–511. https://doi.org/10.1038/35000501.

Almeida, V., L. Schertel Mendes, and D. Doneda. 2023. On the development of AI governance frameworks. *IEEE Internet Computing* 27(1):70–74. https://doi.org/10.1109/MIC.2022.3186030.

Alobayli, F., S. O'Conner, A. Holloway, and K. Cresswell. 2023. Electronic health record stress and burnout among clinicians in hospital settings: A systematic review. *Digital Health* 9. https://doi.org/10.1177%2F20552076231220241.

American Hospital Association (AHA). 2003. *The Patient Care Partnership*. https://www.aha.org/system/files/2018-01/aha-patient-care-partnership.pdf (accessed April 4, 2025).

AHA. 2019. *Surveying the AI Landscape*. https://www.aha.org/system/files/media/file/2019/10/Market_Insights_AI-Landscape.pdf (accessed July 8, 2024).

Anderson, J. G. 2007. Social, ethical and legal barriers to e-health. *International Journal of Medical Informatics* 76(5–6):480–483. https://doi.org/10.1016/j.ijmedinf.2006.09.016.

Anderson, J., and L. Rainie. 2023. *As AI Spreads, Experts Predict the Best and Worst Changes in Digital Life by 2035*. Pew Research Center.

Angwin, J. 2023. Autonomous vehicles are driving blind. *The New York Times*. https://www.nytimes.com/2023/10/11/opinion/driverless-cars-san-francisco.html?smid=url-share (accessed April 4, 2025).

Aquino, Y. S. J., W. A. Rogers, A. Braunack-Mayer, H. Frazer, K. T. Win, N. Houssami, C. Degeling, C. Semsarian, and S. M. Carter. 2023. Utopia versus dystopia: Professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills. *International Journal of Medical Informatics* 169:104903. https://doi.org/10.1016/j.ijmedinf.2022.104903.

Arnaout, A., M. Oseguera-Arasmou, N. Mishra, B. M. Liu, A. Bhattacharya, and D. C. Rhew. 2023. Leveraging technology in public–private partnerships: A model to address public health inequities. *Frontiers in Health Services* 3:1187306. https://doi.org/10.3389/frhs.2023.1187306.

Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD). n.d. *AIM-AHEAD*. https://www.aim-ahead.net (accessed April 4, 2025).

Ash, J. S. 2004. Some unintended consequences of information technology in health care: The nature of patient care information system-related errors. *Journal of the American Medical Informatics Association* 11(2):104–112. https://doi.org/10.1197/jamia.M1471.

Ashton, M. 2018. Getting rid of stupid stuff. *New England Journal of Medicine* 379:1789–1791. https://doi.org/10.1056/NEJMp1809698.

Assistant Secretary for Technology Policy, Office of the National Coordinator for Health IT (ASTP ONC). 2024. *Vulcan Interoperability Bridge Event to Enable*

Clinical Research Infrastructure Information Session. https://www.healthit.gov/news/events/vulcan-interoperability-bridge-event-enable-clinical-research-infrastructure (accessed April 4, 2025).

ASTP ONC. n.d. *Regional Extension Centers (RECs)*. https://www.healthit.gov/topic/regional-extension-centers-recs (accessed April 4, 2025).

Atherton, J. 2011. Development of the electronic health record. *Virtual Mentor* 13(3):186–189. https://doi.org/10.1001/virtualmentor.2011.13.3.mhst1-1103.

Attard-Frost, B., and D. G. Widder. 2024. The ethics of AI value chains. *arXiv*. https://doi.org/10.48550/arXiv.2307.16787.

Awad, E., S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J. F. Bonnefon, and I. Rahwan. 2018. The moral machine experiment. *Nature* 563:59–64. https://doi.org/10.1038/s41586-018-0637-6.

Ayers, J. W., A. Poliak, M. Dredze, E. C. Leas, Z. Zhu, J. B. Kelley, D. J. Faix, A. M. Goodman, C. A. Longhurst, M. l. Hogarth, and D. M. Smith 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine* 183(6):589–596. https://doi.org/10.1001/jamainternmed.2023.1838.

Bajwa, J., U. Munir, A. Nori, and B. Williams. 2021. Artificial intelligence in health care: Transforming the practice of medicine. *Future Healthcare Journal* 8(2):188–194. https://doi.org/10.7861/fhj.2021-0095.

Band, S., A. Yarahmadi, C. Hsu, M. Biyari, M. Sookhak, R. Ameri, I. Dehzangi, A. T. Chronopoulos, and H. Liang. 2023. Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods. *Informatics in Medicine Unlocked* 40:101286. https://doi.org/10.1016/j.imu.2023.101286.

Bandura, A. 2006. Toward a psychology of human agency. *Perspectives on Psychological Science* 1(2):164–180. https://doi.org/10.1111/j.1745-6916.2006.00011.x.

Banja, J., R. D. Hollstein, and M. A. Bruno. 2022. When artificial intelligence models surpass physician performance: Medical malpractice liability in an era of advanced artificial intelligence. *Journal of the American College of Radiology* 19(7):816–820. https://doi.org/10.1016/j.jacr.2021.11.014.

Bates, D. W., and A. Gawande. 2003. Improving safety with information technology. *New England Journal of Medicine* 348:2526–2534. https://doi.org/10.1056/NEJMsa020847.

Bates, D. W., and H. Singh. 2018. Two decades since *To Err Is Human*: An assessment of progress and emerging priorities in patient safety. *Health Affairs* 37(11):1736–1743. https://doi.org/10.1377/hlthaff.2018.0738.

Bates, D. W., L. L. Leape, D. J. Cullen, N. Laird, L. A. Petersen, J. M. Teich, E. Burdick, M. Hickey, S. Kleefield, B. Shea, M. Vander Vliet, and D. L. Seger. 1998.

Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 280(15):1311–1316. https://doi.org/10.1001/jama.280.15.1311.

Bates, D. W., D. Levine, A. Syrowatka, M. Kuznetsova, K. J. T. Craig, A. Rui, G. P. Jackson, and K. Rhee. 2021. The potential of artificial intelligence to improve patient safety: A scoping review. *NPJ Digital Medicine* 4(1):54. https://doi.org/10.1038/s41746-021-00423-6.

Bayati, M., M. Braverman, M. Gillam, K. M. Mack, G. Ruiz, M. S. Smith, and E. Horvitz. 2014. Data-driven decisions for reducing readmissions for heart failure: General methodology and case study. *PLoS ONE* 9(10):e109264. https://doi.org/10.1371/journal.pone.0109264.

Bedoya, A. D., N. J. Economou-Zavlanos, B. A. Goldstein, A. Young, J. E. Jelovsek, C. O'Brien, A. B. Parrish, S. Elengold, K. Lytle, S. Balu, E. Huang, E. G. Poon, and M. J. Pencina. 2022. A framework for the oversight and local deployment of safe and high-quality prediction models. *Journal of the American Medical Informatics Association* 29(9):1631–1636. https://doi.org/10.1093/jamia/ocac078.

Berg, M. 2001. Implementing information systems in health care organizations: Myths and challenges. *International Journal of Medical Informatics* 64(2–3):143–156. https://doi.org/10.1016/S1386-5056(01)00200-3.

Best, E., P. Robles, and D. J. Mallinson. 2024. The future of AI politics, policy and business. *Business and Politics* 26(2):171–179. https://doi.org/10.1017/bap.2024.6.

Birkhäuer, J. 2017. Trust in the health care professional and health outcome: A meta-analysis. *PLoS ONE* 12(2):e0170988. https://doi.org/10.1371/journal.pone.0170988.

Bitterman, D., H. Aerts, and R. Mak. 2020. Approaching autonomy in medical artificial intelligence. *Lancet, Digital Health* 2(9):e447–e449. https://doi.org/10.1016/S2589-7500(20)30187-4.

Blumenthal, D., and B. Patel. 2024. The regulation of clinical artificial intelligence. *NEJM AI* 1(8). https://doi.org/10.1056/AIpc2400545.

Blumenthal, D., and M. Tavenner. 2010. The "meaningful use" regulation for electronic health records. *New England Journal of Medicine* 363(6):501–504. https://doi.org/10.1056/NEJMp1006114.

Borna, S., M. J. Maniaci, C. R. Haider, C. A. Gomez-Cabello, S. M. Pressman, S. A. Haider, B. M. Demaerschalk, J. B. Cowart, and A. J. Forte. 2024. Artificial intelligence support for informal patient caregivers: A systematic review. *Bioengineering* 11(5):483. https://doi.org/10.3390/bioengineering11050483.

Boskma, A., K. van der Braak, N. Ansari, L. Hooft, G. Wietasch, A. Franx, and M. van der Laan. 2023. Assessing the well-being at work of nurses and doctors in hospitals: Protocol for a scoping review of monitoring instruments. *JMIR Research Protocols* 12:e43692. https://doi.org/10.2196%2F43692.

Bouderhem, R. 2024. Shaping the future of AI in health care through ethics and governance. *Humanities and Social Science Communications* 11(416). https://doi.org/10.1057/s41599-024-02894-w.

Bowes, W. A. 2014. Impacts of EHR certification and meaningful use implementation on an integrated delivery network. *AMIA Annual Symposium Proceedings* 2014:325–332.

Brady, A. P., B. Allen, J. Chong, E. Kotter, N. Kottler, J. Mongan, L. Oakden-Rayner, D. Pinto dos Santos, A. Tang, C. Wald, and J. Slavotinek. 2024. Developing, purchasing, implementing and monitoring AI tools in radiology: Practical considerations. A multi-society statement from the ACR, CAR, ESR, RANZCR & RSNA. *Canadian Association of Radiologists Journal* 75(2):226–244. https://doi.org/10.1177/08465371231222229.

Braithwaite, J., P. Glasziou, and J. Westbrook. 2020. The three numbers you need to know about health care: The 60-30-10 Challenge. *BMC Medicine* 18:102. https://doi.org/10.1186/s12916-020-01563-4.

Bremmer, I., and M. Suleyman. 2023. *Building Blocks for AI Governance. International Monetary Fund*. https://www.imf.org/en/Publications/fandd/issues/2023/12/POV-building-blocks-for-AI-governance-Bremmer-Suleyman (accessed April 4, 2025).

Brereton, T. A., M. M. Malik, M. Lifson, J. D. Greenwood, K. J. Peterson, and S. M. Overgaard. 2023. The role of artificial intelligence model documentation in translational science: Scoping review. *Interactive Journal of Medical Research* 12:e45903. https://doi.org/10.2196/45903.

Brown, J. T., Z. Wan, A. Gkoulalas-Divanis, M. Kantarcioglu, and B. A. Malin. 2023. Supporting COVID-19 disparity investigations with dynamically adjusting case reporting policies. *AMIA Annual Symposium Proceedings* 2022:279–288.

Brown, J. A., B. D. Taffe, J. A. Richmond, and M. L. Roberson. 2024. Racial discrimination and health-care system trust among American adults with and without cancer. *Journal of the National Cancer Institute* 116(11):1845–1855. https://doi.org/10.1093/jnci/djae154.

Buchanan, B. G., and E. H. Shortliffe. 1984. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison Wesley.

Budd, J. 2023. Burnout related to electronic health record use in primary care. *Journal of Primary Care & Community Health* 14. https://doi.org/10.1177%2F21501319231166921.

Buocz, T., S. Pfotenhauer, and I. Eisenberger. 2023. Regulatory sandboxes in the AI Act: Reconciling innovation and safety? *Law, Innovation and Technology* 15(2):357–389. https://doi.org/10.1080/17579961.2023.2245678.

Caldwell, C., J. Brexler, and T. Gillem. 2005. *Lean-Six Sigma for Healthcare: A Senior Leader's Guide to Improving Cost and Throughput*. American Society for Quality.

Carnegie Mellon University Software Engineering Institute. 2021. *National AI Engineering Initiative: Scalable AI*. https://insights.sei.cmu.edu/documents/608/2021_019_001_735330.pdf (accessed July 8, 2024).

Carroll, S. R., E. Herczog, M. Hudson, K. Russell, and S. Stall. 2021. Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data* 8(1):108. https://doi.org/10.1038/s41597-021-00892-0.

Cary, M. P., A. Zink, S. Wei, A. Olson, M. Yan, R. Senior, S. Bessias, K. Gadhoumi, G. Jean-Pierre, D. Wang, L. S. Ledbetter, N. J. Economou-Zavlanos, Z. Obermeyer, and M. J. Pencina. 2023. Mitigating racial and ethnic bias and advancing health equity in clinical algorithms: A scoping review. *Health Affairs* 42(10):1359–1368. https://doi.org/10.1377/hlthaff.2023.00553.

Centers for Disease Control and Prevention (CDC). 2021. *NEHRS Public Use File National Weighted Estimates.* https://www.cdc.gov/nchs/data/nehrs/2021NEHRS-PUF-weighted-estimates-508.pdf (accessed April 4, 2025).

CDC. 2022. *Looking at AI's Potential Impact on Health Equity*. https://www.cdc.gov/surveillance/data-modernization/snapshot/2022-snapshot/stories/ai-impact-health-equity.html (accessed July 3, 2024).

CDC. 2024. *National Diabetes Statistics Report*. https://www.cdc.gov/diabetes/php/data-research/index.html (accessed July 13, 2024).

CDC. n.d. *The Untreated Syphilis Study at Tuskegee Timeline*. https://www.cdc.gov/tuskegee/timeline.htm (accessed July 30, 2024).

Centers for Medicare & Medicaid Services (CMS). 2024. *National Health Expenditure Fact Sheet*. https://www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data/nhe-fact-sheet (accessed April 16, 2025).

Chassin, M. R., and R. W. Galvin. 1998. The urgent need to improve health care quality. Institute of Medicine National Roundtable on Health Care Quality. *JAMA* 280(11):1000–1005. https://doi.org/10.1001/jama.280.11.1000.

Chaudhry, B., J. Wang, and S. Wu. 2006. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine* 144(10):742–752. https://doi.org/10.7326/0003-4819-144-10-200605160-00125.

Chen, H., C. Gomez, C. M. Huang, and M. Unberath. 2022. Explainable medical imaging AI needs human-centered design: Guidelines and evidence from a systematic review. *NPJ Digital Medicine* 5:156. https://doi.org/10.1038/s41746-022-00699-2.

Chen, M., L. Parks Golding, and G. N. Nicla. 2021. Who will pay for AI? *Radiology Artificial Intelligence* 3(3). https://doi.org/10.1148/ryai.2021210030.

Christiano, A., and A. Neimand. 2018. The science of what makes people care. *Stanford Social Innovation Review* 16(4):26–33. https://doi.org/10.48558/GW2V-5279.

Coalition for Health AI (CHAI). n.d. *Our Workstreams*. https://chai.org/workstreams (accessed April 4, 2025).

Cohen, M. F. 2016. Impact of the HITECH financial incentives on EHR adoption in small, physician-owned practices. *International Journal of Medical Informatics* 94:143–154. https://doi.org/10.1016/j.ijmedinf.2016.06.017.

Coiera, E. W., K. Verspoor, and D. P. Hansen. 2023. We need to chat about artificial intelligence. *Medical Journal of Australia* 219(3):98–100. https://doi.org/10.5694/mja2.51992.

Compton-Phillips, A. 2019. Spreading at scale: A practical leadership model for change. *NEJM Catalyst* 1(1). https://doi.org/10.1056/CAT.19.1083.

Consumer Technology Association. 2023. *Artificial Intelligence in Health Care: Practices for Identifying and Managing Bias (ANSI/CTA-2116)*. https://shop.cta.tech/collections/standards/products/artificial-intelligence-in-health-care-practices-for-identifying-and-managing-bias-cta-2116 (accessed July 3, 2024).

Crafts, N. 2021. Artificial intelligence as a general-purpose technology: An historical perspective. *Oxford Review of Economic Policy* 37(3):521–536. https://doi.org/10.1093/oxrep/grab012.

Cutler, D. 2023. What artificial intelligence means for health care. *JAMA Health Forum* 4(7):e232652. https://doi.org/10.1001/jamahealthforum.2023.2652.

Dafoe, A. 2018. *AI Governance: A Research Agenda.* Future of Humanity Institute. https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf (accessed April 4, 2025).

Damiani, G., G. Altamura, M. Zedda, M. C. Nurchis, G. Aulino, A. Heidar Alizadeh, F. Cazzato, G. Della Morte, M. Caputo, S. Grassi, A. Oliva, and D.3.2 Group. 2023. Potentiality of algorithms and artificial intelligence adoption to improve medication management in primary care: A systematic review. *BMJ Open* 13(3):e065301. https://doi.org/10.1136/bmjopen-2022-065301.

Daneshjou, R., M. P. Smith, M. D. Sun, V. Rotemberg, and J. Zou. 2021. Lack of transparency and potential bias in artificial intelligence data sets and algorithms: A scoping review. *JAMA Dermatology* 157:1362–1369. https://doi.org/10.1001/jamadermatol.2021.3129.

Data Science Process Alliance. n.d. *The GenAI Lifecycle*. https://www.datascience-pm.com/the-genai-life-cycle (accessed April 4, 2025).

Davenport, T., and R. Kalakota. 2019. The potential for artificial intelligence in healthcare. *Future Healthcare* 6(2):94–98. https://doi.org/10.7861/futurehosp.6-2-94.

Davis, S. E., T. A. Lasko, G. Chen, E. D. Siew, and M. E. Matheny. 2017. Calibration drift in regression and machine learning models for acute kidney injury. *Journal of the American Medical Informatics Association* 24(6):1052–1061. https://doi.org/10.1093/jamia/ocx030.

Davis, S. E., R. A. Greevy, Jr., T. A. Lasko, C. G. Walsh, and M. E. Matheny. 2020. Detection of calibration drift in clinical prediction models to inform model updating. *Journal of Biomedical Informatics* 112:103611. https://doi.org/10.1016/j.jbi.2020.103611.

Davis, S. E., C. G. Walsh, and M. E. Matheny. 2022. Open questions and research gaps for monitoring and updating AI-enabled tools in clinical settings. *Frontiers in Digital Health* 4:958284. https://doi.org/10.3389/fdgth.2022.958284.

Davis, S. E., P. J. Embí, and M. E. Matheny. 2024. Sustainable deployment of clinical prediction tools—a 360° approach to model maintenance. *Journal of the American Medical Informatics Association* 31(5):1195–1198. https://doi.org/10.1093/jamia/ocae036.

de Man, Y., Y Wieland-Jorna, B. Torensma, K. de Wit, A. L. Francke, M. G. Oosterveld-Vlug, and R. A. Verheij. 2023. Opt-in and opt-out consent procedures for the reuse of routinely recorded health data in scientific research and their consequences for consent rate and consent bias: Systematic review. *Journal of Medical Internet Research* 25:e42131. https://doi.org/10.2196%2F42131.

De Silva, D., and D. Alahakoon. 2022. An artificial intelligence life cycle: From conception to production. *Patterns* 3(6):100489. https://doi.org/10.1016/j.patter.2022.100489.

de Vries, N., A. Boone, L. Godderis, J. Bouman, S. Szemik, D. Matranga, and P. de Winter. 2023. The race to retain healthcare workers: A systematic review on factors that impact retention of nurses and physicians in hospitals. *Inquiry* 60:469580231159318. https://doi.org/10.1177%2F00469580231159318.

Dean, J. 2022. A golden decade of deep learning: Computing systems & applications. *Daedalus* 151(2):58–74. https://doi.org/10.1162/daed_a_01900.

Department of Health and Human Services (HHS). 1979. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html (accessed July 31, 2024).

HHS. 1996. *Health Insurance Portability and Accountability Act of 1996*. https://aspe.hhs.gov/reports/health-insurance-portability-accountability-act-1996 (accessed April 4, 2025).

HHS. 2009. *Health Information Technology for Economic and Clinical Health Act* (HITECH Act). https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/hitechact.pdf (accessed August 2, 2024).

HHS. 2014. *Effects of Meaningful Use Functionalities on Health Care Quality, Safety and Efficiency.* https://dashboard.healthit.gov/quickstats/pages/FIG-Health-IT-Literature-Review-Summary-of-Author-Sentiments.php (accessed April 4, 2025).

HHS. 2021. *Trustworthy AI (TAI) Playbook*. https://www.hhs.gov/sites/default/files/hhs-trustworthy-ai-playbook.pdf (accessed July 3, 2024).

HHS. 2023a. *HHS Finalizes Rule to Advance Health IT Interoperability and Algorithm Transparency*. https://www.hhs.gov/about/news/2023/12/13/hhs-finalizes-rule-to-advance-health-it-interoperability-and-algorithm-transparency.html (accessed July 3, 2024).

HHS. 2023b. *Our Epidemic of Loneliness and Isolation*. https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf (accessed April 4, 2025).

HHS. 2024a. *HHS Issues New Rule to Strengthen Nondiscrimination Protections and Advance Civil Rights in Health Care*. https://www.hhs.gov/about/news/2024/04/26/hhs-issues-new-rule-strengthen-nondiscrimination-protections-advance-civil-rights-health-care.html (accessed July 14, 2024).

HHS. 2024b. *HHS Reorganizes Technology, Cybersecurity, Data, and Artificial Intelligence Strategy and Policy Functions*. https://www.hhs.gov/about/news/2024/07/25/hhs-reorganizes-technology-cybersecurity-data-artificial-intelligence-strategy-policy-functions.html (accessed April 4, 2025).

HHS. 2024c. Individuals' right under HIPAA to access their health information, 45 CFR § 164.524. https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/index.html (accessed April 4, 2025).

HHS. 2025. *U.S. Department of Health and Human Services: Strategic Plan for the Use of Artificial Intelligence in Health, Human Services, and Public Health*. https://www.healthit.gov/sites/default/files/2025-01/2025%20HHS%20AI%20Strategic%20Plan_Full_508.pdf (accessed January 13, 2025).

HHS. n.d.a. *HHS Artificial Intelligence (AI) Strategy*. https://www.hhs.gov/programs/topic-sites/ai/strategy/index.html (accessed July 3, 2024).

HHS. n.d.b. *About HHS*. https://www.hhs.gov/about/index.html (accessed July 3, 2024).

Department of Justice. n.d. *Title VI of the Civil Rights Act of 1964*. https://www.justice.gov/crt/fcs/TitleVI (accessed April 4, 2025).

Dorr, D. A., L. Adams, and P. Embí. 2023. Harnessing the promise of artificial intelligence responsibly. *JAMA* 329(16):1347–1348. https://doi.org/10.1001/jama.2023.2771.

du Preez, A., S. Bhattacharya, P. Beling, and E. Bowen. 2025. Fraud detection in healthcare claims using machine learning: A systematic review. *Artificial Intelligence in Medicine* 160:103061. https://doi.org/10.1016/j.artmed.2024.103061.

Duke AI Health. n.d. *What Is ABCDS?* https://aihealth.duke.edu/algorithm-based-clinical-decision-support-abcds (accessed April 4, 2025).

Echo Wang, H., M. Landers, R. Adams, A. Subbaswamy, H. Kharrazi, D. Gaskin, and S. Saria. 2022. A bias evaluation checklist for predictive models and its pilot application for 30-day hospital readmission models. *Journal of the American Medical Informatics Association* 29(8):1323–1333. https://doi.org/10.1093/jamia/ocac065.

Eldridge, N., Y. Wang, M. Metersky, S. Eckenrode, J. Mathew, N. Sonnenfeld, J. Perdue-Puli, D. Hunt, P. J. Brady, P. McGann, E. Grace, D. Rodrick, E. Drye, and H. M. Krumholz. 2022. Trends in adverse event rates in hospitalized patients, 2010–2019. *JAMA* 328(2):173–183. https://doi.org/10.1001/jama.2022.9600.

Emanuel, L., D. Berwick, J. Conway, J. Combes, M. Hatlie, L. Leape, J. Reason, P. Schyve, C. Vincent, and M. Walton. 2008. What exactly is patient safety? In *Advances in Patient Safety: New Directions and Alternative Approaches,* Vol. 1, edited by K. Henriksen, J. B. Battles, M. A. Keyes, and M. L. Grady. Rockville, MD: Agency for Healthcare Research and Quality.

Embí, P. J. 2021. Algorithmovigilance—advancing methods to analyze and monitor artificial intelligence–driven health care for effectiveness and equity. *JAMA Network Open* 4(4):e214622. https://doi.org/10.1001/jamanetworkopen.2021.4622.

Escobar, G. J., V. X. Liu, A. Schuler, B. Lawson, J. D. Greene, and P. Kipnis. 2020. Automated identification of adults at risk for in-hospital clinical deterioration. *New England Journal of Medicine* 383(20):1951–1960. https://doi.org/10.1056/NEJMsa2001090.

Esmaeilzadeh, P., T. Mirzaei, and S. Dharanikota. 2021. Patients' perceptions toward human–artificial intelligence interaction in health care: Experimental study. *Journal of Medical Internet Research* 23(11):e25856. https://doi.org/10.2196/25856.

Esteva, A., A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean. 2019. A guide to deep learning in healthcare. *Nature Medicine* 25(1):24–29. https://doi.org/10.1038/s41591-018-0316-z.

European Commission. 2019. *Ethics Guidelines for Trustworthy AI.* https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed April 4, 2025).

European Commission. 2024a. *AI for Public Good: EU-U.S. Research Alliance in AI for the Public Good*. https://digital-strategy.ec.europa.eu/en/library/ai-public-good-eu-us-research-alliance-ai-public-good (accessed April 4, 2025).

European Commission. 2024b. *EU and US Continue Strong Trade and Technology Cooperation at a Time of Global Challenges.* https://ec.europa.eu/commission/presscorner/detail/en/IP_24_1827 (accessed April 4, 2025).

European Medicines Agency (EMA). 2021. *Guideline on Clinical Evaluation of Medical Devices*. https://www.ema.europa.eu/en/human-regulatory-overview/medical-devices (accessed April 4, 2025).

European Observatory on Health Systems and Policies. 2016. *Strengthening Health System Governance: Better Policies, Stronger Performance.* Open University Press.

Farrar, B., G. Wang, H. Bos, D. Schneider, H. Noel, J. Guo, L. Koester, A. Desai, K. Manson, S. Garfinkel, A. Ptaszek, and M. Dalldorf. 2015. *Evaluation of the Regional Extension Center Program: Final Report*. Washington, DC: Office of the National Coordinator for Health Information Technology.

Fawzy, A., T. D. Wu, K. Wang, M. L. Robinson, J. Farha, A. Bradke, S. H. Golden, Y. Xu, and B. T. Garibaldi. 2022. Racial and ethnic discrepancy in pulse oximetry and delayed identification of treatment eligibility among patients with COVID-19. *JAMA Internal Medicine* 182(7):730–738. https://doi.org/10.1001/jamainternmed.2022.1906.

Fehr, J., B. Citro, R. Malpani, C. Lippert, and V. I. Madai. 2024. A trustworthy AI reality-check: The lack of transparency of artificial intelligence products in healthcare. *Frontiers in Digital Health* 6:1267290. https://doi.org/10.3389/fdgth.2024.1267290.

Feng, J., R. V. Phillips, I. Malenica, A. Bishara, A. E. Hubbard, L. A. Celi, and R. Pirracchio. 2022. Clinical artificial intelligence quality improvement: Towards continual monitoring and updating of AI algorithms in health care. *NPJ Digital Medicine* 5(1):66. https://doi.org/10.1038/s41746-022-00611-y.

Fernandes, M. S., and J. R. Goldim. 2024. Artificial intelligence and decision making in health: Risks and opportunities. In *Multidisciplinary Perspectives on Artificial Intelligence and the Law. Law, Governance and Technology Series*, Vol. 58, edited by H. Sousa Antunes, P. M. Freitas, A. L. Oliveira, C. Martins Pereira, E. Vaz de Sequeira, and L. Barreto Xavier. Cham: Springer, pp. 187–205. https://doi.org/10.1007/978-3-031-41264-6_10.

Ferrara, E. 2023. Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Science* 6(1):3. https://doi.org/10.3390/sci6010003.

Finlayson, S. G., A. Subbaswamy, K. Singh, J. Bowers, A. Kupke, J. Zittrain, I. S. Kohane, and S. Saria. 2021. The clinician and dataset shift in artificial intelligence. *New England Journal of Medicine* 385(3):283–286. https://doi.org/10.1056/NEJMc2104626.

Fleisher, L. A., and N. J. Economou-Zavlanos. 2024. Artificial intelligence can be regulated using current patient safety procedures and infrastructure in hospitals. *JAMA Health Forum* 5(6):e241369. https://doi.org/10.1001/jamahealthforum.2024.1369.

Fleisher, L. A., M. Schreiber, D. Cardo, and A. Srinivasan. 2022. Health care safety during the pandemic and beyond—building a system that ensures resilience. *New England Journal of Medicine* 386(7):609–611. https://doi.org/10.1056%2FNEJMp2118285.

Fuggetta, A. 2000. Software process: A roadmap. Presented at the Conference on the Future of Software Engineering, Limerick, Ireland.

Furukawa, M. F., J. King, and V. Patel. 2015. Physician attitudes on ease of use of EHR functionalities related to meaningful use. *American Journal of Managed Care* 21(12):e684–e692.

Gartner. n.d. *Gartner Hype Cycle.* https://www.gartner.com/en/research/methodologies/gartner-hype-cycle?utm_source (accessed January 13, 2025).

Gianni, R., S. Lehtinen, and M. Nieminen. 2022. Governance of responsible AI: From ethical guidelines to cooperative policies. *Frontiers in Computer Science* 4. https://doi.org/10.3389/fcomp.2022.873437.

Glauberman, G., A. Ito-Fujita, S. Katz, and J. Callahan. 2023. Artificial intelligence in nursing education: Opportunities and challenges. *Hawaii Journal of Health and Social Welfare* 82(12):302–305.

Gomes, Y. E., M. Chau, H. A. Banwell, and R. S. Causby. 2022. Diagnostic accuracy of the Ottawa ankle rule to exclude fractures in acute ankle injuries in adults: A systematic review and meta-analysis. *BMC Musculoskeletal Disorders* 23:885. https://doi.org/10.1186/s12891-022-05831-7.

Gonzales, A., G. Guruswamy, and S. R. Smith. 2023. Synthetic data in health care: A narrative review. *PLoS Digital Health* 2(1):e0000082. https://doi.org/10.1371/journal.pdig.0000082.

Greeley, C., L. Holder, E. E. Nilsson, and M. K. Skinner. 2024. Scalable deep learning artificial intelligence histopathology slide analysis and validation. *Scientific Reports* 14:26748. https://doi.org/10.1038/s41598-024-76807-x.

Green, L. A., G. Potworowski, A. Day, R. May-Gentile, D. Vibbert, B. Maki, and L. Kiesel. 2015. Sustaining "meaningful use" of health information technology in low-resource practices. *Annals of Family Medicine* 13(1):17–22. https://doi.org/10.1370/afm.1740.

Greenhalgh, T., J. Wherton, C. Papoutsi, J. Lynch, G. Hughes, C. A'Court, S. Hinder, N. Fahy, R. Procter, and S. Shaw. 2017. Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *Journal of Medical Internet Research* 19(11):e367. https://doi.org/10.2196/jmir.8775.

Grother, P., M. Ngan, and K. Hanaoka. 2019. *Face Recognition Vendor Test (FRVT) Part 3* (NISTIR 8280). National Institute of Standards and Technology. http://dx.doi.org/10.6028/NIST.IR.8280.

Guillaudeux, M., O. Rousseau, J. Petot, Z. Bennis, C.-A. Dein, T. Goronflot, N. Vince, S. Limou, M. Karakachoff, M. Wargny, and P.-A. Gourraud. 2023. Patient-centric synthetic data generation, no reason to risk re-identification in biomedical data analysis. *NPJ Digital Medicine* 6(1):37. https://doi.org/10.1038/s41746-023-00771-5.

Gulshan, V., L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster. 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316(22):2402–2410. https://doi.org/10.1001/jama.2016.17216.

Habib, A., A. Lin, and R. W. Grant. 2021. The Epic Sepsis Model falls short—the importance of external validation. *JAMA Internal Medicine* 181(8):1040–1041. https://doi.org/10.1001/jamainternmed.2021.3333.

Habicht, J., S. Viswanathan, B. Carrington, T. U. Hauser, R. Harper, and M. Rollwage. 2024. Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot. *Nature Medicine* 30(2):595–602. https://doi.org/10.1038/s41591-023-02766-x.

Halamka, J., and P. Cerrato. 2021. Understanding the role of digital platforms in technology readiness. *Regenerative Medicine* 16(3):207–213. https://doi.org/10.2217/rme-2020-0135.

Halamka, J., and M. Tripathi. 2017. The HITECH era in retrospect. *New England Journal of Medicine* 377(10):907–909. https://doi.org/10.1056/NEJMp1709851.

Han, Y. Y., J. Carcillo, S. Venkataraman, R. S. B. Clark, R. S. Watson, T. C. Nguyen, H. Bayir, and R. A. Orr. 2005. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 116(6):1506–1512.

Haneuse, S. 2016. Distinguishing selection bias and confounding bias in comparative effectiveness research. *Medical Care* 54(4):e23–e29. https://doi.org/10.1097/mlr.0000000000000011.

Health AI Partnership (HAIP). n.d. *Key Decision Points.* https://healthaipartnership.org/key-decisions-in-adopting-an-ai-solution (accessed April 4, 2025).

Heckerman, D. E., E. J. Horvitz, and B. N. Nathwani. 1992. Toward normative expert systems: Part I. The Pathfinder project. *Methods of Information in Medicine* 31(2):90–105.

Heilinger, J. C. 2022. The ethics of AI ethics. A constructive critique. *Philosophy & Technology* 35:61. https://doi.org/10.1007/s13347-022-00557-9.

Henry, K. E., R. Kornfield, A. Sridharan, R. C. Linton, C. Groh, T. Wang, A. Wu, B. Mutlu, and S. Saria. 2022. Human–machine teaming is key to AI adoption: Clinicians' experiences with a deployed machine learning system. *NPJ Digital Medicine* 5(1):97. https://doi.org/10.1038/s41746-022-00597-7.

Hicks, S. A., I. Strümke, V. Thambawita, M. Hammou, M. A. Riegler, P. Halvorsen, and S. Parasa. 2022. On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports* 12:5979. https://doi.org/10.1038/s41598-022-09954-8.

Hiefner, A. R., P. Constable, K. Ross, D. Sepdham, and J. B. Ventimiglia. 2022. Protecting family physicians from burnout: Meaningful patient-physician relationships are "more than just medicine." *Journal of the American Board of Family Medicine* 35(4):716–723. https://doi.org/10.3122/jabfm.2022.04.210441.

Hillestad, R., J. Bigelow, A. Bower, F. Girosi, R. Meili, R. Scoville, and R. Taylor. 2005. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Affairs (Millwood)* 24(5):1103–1117. https://doi.org/10.1377/hlthaff.24.5.1103.

Hoff, T., K. Trovato, and A. Kitsakos. 2023. Burnout among family physicians in the United States: A review of the literature. *Quality Management in Health Care* 33(1):1–11. https://doi.org/10.1097/qmh.0000000000000439.

Hogg, H. D. J., M. Al-Zubaidy, Technology Enhanced Macular Services Study Reference Group, J. Talks, A. K. Denniston, C. J. Kelly, J. Malawana, C. Papoutsi, M. D. Teare, P. A. Keane, F. R. Beyer, and G. Maniatopoulos. 2023. Stakeholder perspectives of clinical artificial intelligence implementation: Systematic review of qualitative evidence. *Journal of Medical Internet Research* 25:e39742. https://doi.org/10.2196/39742.

Holmgren, A. J., N. L. Downing, D. W. Bates, T. D. Shanafelt, A. Milstein, C. D. Sharp, D. M. Cutler, R. S. Huckman, and K. A. Schulman. 2021. Assessment of electronic health record use between US and non-US health systems. *JAMA Internal Medicine* 181(2):251–259. https://doi.org/10.1001/jamainternmed.2020.7071.

Holmgren, A. J., N. Hendrix, N. Maisel, J. Everson, A. Bazemore, L. Rotenstein, R. L. Phillips, and J. Adler-Milstein. 2024. Electronic health record usability, satisfaction, and burnout for family physicians. *JAMA Network Open* 7(8):e2426956. https://doi.org/10.1001/jamanetworkopen.2024.26956.

Hong, G. S., M. Jang, S. Kyung, K. Cho, J. Jeong, G. Y. Lee, K. Shin, K. D. Kim, S. M. Ryu, J. B. Seo, S. M. Lee, and N. Kim. 2023. Overcoming the challenges in the development and implementation of artificial intelligence in radiology: A comprehensive review of solutions beyond supervised learning. *Korean Journal of Radiology* 24(11):1061–1080. https://doi.org/10.3348/kjr.2023.0393.

Horvitz, E., and T. M. Mitchell. 2024. Scientific progress in artificial intelligence: History, status, and futures. In *Realizing the Promise and Minimizing the Perils of AI for Science and the Scientific Community*, edited by K. Hall Jamieson, A.-M. Mazza, and W. Kearney. University of Pennsylvania Press.

Horvitz, E. J., J. S. Breese, and M. Henrion. 1988. Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning* 2(3):247–302. https://doi.org/10.1016/0888-613X(88)90120-X.

Householder, A., G. Wassermann, A. Manion, and C. King. 2017. *The CERT Guide to Coordinated Vulnerability Disclosure* (Special Report CMU/SEI-2017-SR-022). Carnegie Mellon University Software Engineering Institute. https://insights.sei.cmu.edu/documents/1945/2017_003_001_503340.pdf (accessed July 3, 2024).

Howard, J. 2022. Algorithms and the future of work. *American Journal of Industrial Medicine* 65(12):943–952. https://doi.org/10.1002/ajim.23429.

Howell, M. D., G. S. Corrado, and K. B. DeSalvo. 2024. Three epochs of artificial intelligence in health care. *JAMA* 331(3):242–244. https://doi.org/10.1001/jama.2023.25057.

Ibrahim, H., X. Liu, S. C. Rivera, D. Moher, A. W. Chan, M. R. Sydes, M. J. Calvert, and A. K. Denniston. 2021. Reporting guidelines for clinical trials of artificial intelligence interventions: The SPIRIT-AI and CONSORT-AI guidelines. *Trials* 22(1):11. https://doi.org/10.1186/s13063-020-04951-6.

Institute of Medicine (IOM). 1991. *Computer-Based Patient Record: An Essential Technology for Health Care*. Washington, DC: National Academy Press. https://doi.org/10.17226/18459.

IOM. 2000. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press.

IOM. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC: National Academy Press.

IOM. 2002. Medical innovation in the changing healthcare marketplace: Conference summary. In *Barriers to Medical Innovation*. Washington, DC: The National Academies Press.

IOM. 2003. *Key Capabilities of an Electronic Health Record System: Letter Report*. Washington, DC: The National Academies Press.

IOM. 2007. *Preventing Medication Errors: Quality Chasm Series*. Washington, DC: The National Academies Press.

IOM. 2011. *Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care*. Washington, DC: The National Academies Press.

International Data Corporation (IDC). 2023. *The Business Opportunity of AI: How Leading Organizations Around the World Are Using AI to Drive Impact Across Every Industry*. https://news.microsoft.com/source/wp-content/uploads/2023/11/US51315823-IG-ADA.pdf (accessed April 4, 2025).

International Organization for Standardization (ISO). 2023. *ISO/IEC 42001*. https://www.iso.org/obp/ui/en/#iso:std:iso-iec:42001:ed-1:v1:en (accessed April 4, 2025).

ISO. n.d. *Health Informatics*. https://www.iso.org/sectors/health/health-informatics?utm_campaign=it-link&utm_source=internal (accessed April 4, 2025).

Jabbour, S., D. Fouhey, S. Shepard, T. S. Valley, E. A. Kazerooni, N. Banovic, J. Wiens, and M. W. Sjoding. 2023. Measuring the impact of AI in the diagnosis of hospitalized patients: A randomized clinical vignette survey study. *JAMA* 330(23):2275–2284. https://doi.org/10.1001/jama.2023.22295.

Jacob, C., N. Brasier, E. Laurenzi, S. Heuss, S. G. Mougiakakou, A. Cöltekin, and M. K. Peter. 2025. AI for IMPACTS framework for evaluating the long-term real-world impacts of AI-powered clinician tools: Systematic review and narrative synthesis. *Journal of Medical Internet Research* 27:e67485. https://doi.org/10.2196/67485.

Jia, Z., J. Chen, X. Xu, J. Kheir, J. Hu, H. Xiao, S. Peng, X. S. Hu, D. Chen, and Y. Shi. 2023. The importance of resource awareness in artificial intelligence for health care. *Nature Machine Intelligence* 5:687–698. https://doi.org/10.1038/s42256-023-00670-0.

Jiang, F., Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang. 2017. Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology* 2(4):230–243. https://doi.org/10.1136/svn-2017-000101.

Johnson, K. B., and W. Stead. 2022. Making electronic health records both SAFER and SMARTER. *JAMA* 328(6):523–524. https://doi.org/10.1001/jama.2022.12243.

Johnson, K. B., M. J. Neuss, and D. E. Detmer. 2021. Electronic health records and clinician burnout: A story of three eras. *Journal of the American Medical Informatics Association* 28(5):967–973. https://doi.org/10.1093/jamia/ocaa274.

Johnson, L. C., R. Beaton, S. Murphy, and K. Pike. 2000. Sampling bias and other methodological threats to the validity of health survey research. *International Journal of Stress Management* 7:247–267. https://doi.org/10.1023/A:1009589812697.

Joseph, S., J. Selvaraj, I. Mani, T. Kumaragurupari, X. Shang, P. Mudgil, T. Ravilla, and M. He. 2024. Diagnostic accuracy of artificial intelligence-based automated diabetic retinopathy screening in real-world settings: A systematic review and meta-analysis. *American Journal of Ophthalmology* 263:214–230. https://doi.org/10.1016/j.ajo.2024.02.012.

Juhn, Y. J., E. Ryu, C. Wi, K. S. King, M. Malik, S. Romero-Brufau, C. Weng, S. Sohn, R. R. Sharp, and J. D. Halamka. 2022. Assessing socioeconomic bias in machine learning algorithms in health care: A case study of the HOUSES index. *Journal of the American Medical Informatics Association* 29(7):1142–1151. https://doi.org/10.1093%2Fjamia%2Focac052.

Jung, K. H. 2023. Uncover this tech term: Foundation model. *Korean Journal of Radiology* 24(10):1038–1041. https://doi.org/10.3348/kjr.2023.0790.

Kaelber, D., P. Greco, and R. D. Cebul. 2005. Evaluation of a commercial electronic medical record (EMR) by primary care physicians 5 years after implementation. *AMIA Annual Symposium Proceedings* 2005:1002.

Kalkman, S., M. Mostert, C. Gerlinger, J. J. M. van Delden, and G. J. M. W. van Thiel. 2019. Responsible data sharing in international health research: A systematic review of principles and norms. *BMC Medical Ethics* 20(1):21. https://doi.org/10.1186/s12910-019-0359-9.

Kalkman, S., J. van Delden, A. Banerjee, B. Tyl, M. Mostert, and G. van Thiel. 2022. Patients' and public views and attitudes towards the sharing of health data for research: A narrative review of the empirical evidence. *Journal of Medical Ethics* 48(1):3–13. https://doi.org/10.1136/medethics-2019-105651.

Kaur, A., A. Budko, K. Liu, E. Eaton, B. Steitz, and K. B. Johnson. 2024. Automating responses to patient portal messages using generative AI. *medRxiv*. https://doi.org/10.1101/2024.04.25.24306183.

Kemmerling, M., D. Lütticke, and R. H. Schmitt. 2024. Beyond games: A systematic review of neural Monte Carlo tree search applications. *Applied Intelligence* 54:1020–1046. https://doi.org/10.1007/s10489-023-05240-w.

Kennedy-Moulton, K., S. Miller, P. Persson, M. Rossin-Slater, L. Wherry, and G. Aldana. 2023. *Maternal and Infant Health Inequality: New Evidence from Linked Administrative Data* (NBER Working Paper 30693). National Bureau of Economic Research. https://doi.org/10.3386/w30693.

Kesavan, P., and C. J. Dy. 2020. Impact of health care reform on technology and innovation. *Hand Clinics* 36(2):255–262. https://doi.org/10.1016/j.hcl.2020.01.008.

Khalid, N., A. Qayyum, M. Bilal, A. Al-Fuqaha, and J. Qadir. 2023. Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine* 158:106848. https://doi.org/10.1016/j.compbiomed.2023.106848.

Khezeli, K., S. Siegel, B. Shickel, T. Ozrazgat-Baslanti, A. Bihorac, and P. Rashidi. 2023. Reinforcement learning for clinical applications. *Clinical Journal of the American Society of Nephrology* 18(4):521–523. https://doi.org/10.2215/CJN.0000000000000084.

Kim, J. Y., W. Boag, F. Gulamali, A. Hasan, H. D. J. Hogg, M. Lifson, D. Mulligan, M. Patel, I. D. Raji, A. Sehgal, K. Shaw, D. Tobey, A. Valladares, D. Vidal, S. Balu, and M. Sendak. 2023. Organizational governance of emerging technologies: AI adoption in healthcare. *arxiv*. https://doi.org/10.48550/arXiv.2304.13081.

Kim, J.Y., A. Hasan, K. C. Kellogg, W. Ratliff, S. G. Murray, H. Suresh, A. Valladares, K. Shaw, D. Tobey, D. E. Vidal, M. A. Lifson, M. Patel, I. Deborah Raji, M. Gao, W. Knechtle, L. Tang, S. Balu, and M. P. Sendak. 2024. Development and preliminary testing of Health Equity Across the AI Lifecycle (HEAAL): A framework for healthcare delivery organizations to mitigate the risk of AI solutions worsening health inequities. *PLoS Digital Health* 3(5):e0000390. https://doi.org/10.1371%2Fjournal.pdig.0000390.

Kiseleva, A., D. Kotzinos, and P. De Hert. 2022. Transparency of AI in healthcare as a multilayered system of accountabilities: Between legal requirements and technical limitations. *Frontiers in Artificial Intelligence* 5:879603. http://dx.doi.org/10.3389/frai.2022.879603.

Knapp, A. N., E. Dow, K. Chen, N. C. Khan, D.V. Do, V. Mahajan, P. Mruthyunjaya, T. Leng, and D. Myung. 2023. Real world outcomes from artificial intelligence to detect diabetic retinopathy in the primary care setting: 12 month experience. *Investigative Ophthalmology & Visual Science* 64(8):252.

Koller, D., A. Beam, A. Manrai, E. Ashley, M. B., X. Liu, J. Gichoya, C. Holmes, J. Zou, N. Dagan, T.Y. Wong, D. Blumenthal, and I. Kohane. 2024. Why we support and encourage the use of large language models in NEJM AI submissions. *NEJM AI* 1(1). https://doi.org/10.1056/AIe2300128.

Komorowski, M., L. A. Celi, O. Badawi, A. C. Gordon, and A. Aldo Faisal. 2018. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine* 24:1716–1720. https://doi.org/10.1038/s41591-018-0213-5.

Kore, A., E. A. Bavil, V. Subasri, M. Abdalla, B. Fine, E. Dolatabadi, and M. Abdalla. 2024. Empirical data drift detection experiments on real-world medical imaging data. *Nature Communications*. 15:1887. https://doi.org/10.1038/s41467-024-46142-w.

Kovacs, R. J. 2015. Using practice-based evidence to reduce disparities in care: A virtuous cycle turns. *Journal of the American College of Cardiology* 66(11):1234–1235. https://doi.org/10.1016/j.jacc.2015.07.046.

Krittanawong, C., and S. Kaplin. 2021. Artificial intelligence in global health. *European Heart Journal* 42(24):2321–2322. https://doi.org/10.1093/eurheartj/ehab036.

Krive, J., M. Isola, L. Chang, T. Patel, M. Anderson, and R. Sreedhar. 2023. Grounded in reality: Artificial intelligence in medical education. *JAMIA Open* 6(2):ooad037. https://doi.org/10.1093/jamiaopen/ooad037.

Kshetri, N. 2024. Economics of artificial intelligence governance. *Computer* 57(4):113–118. https://doi.org/10.1109/MC.2024.3357951.

Kumar, Y., A. Koul, R. Singla, and M. F. Ijaz. 2023. Artificial intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda. *Journal of Ambient Intelligence and Humanized Computing* 14:8459–8486. https://doi.org/10.1007/s12652-021-03612-z.

Kuo, R. Y. L., A. Freethy, J. Smith, R. Hill, J. C, D. Jerome, E. Harriss, G. S. Collins, E. Tutton, and D. Furniss. 2024. Stakeholder perspectives towards diagnostic artificial intelligence: A co-produced qualitative evidence synthesis. *EClinicalMedicine* 71:102555. https://doi.org/10.1016%2Fj.eclinm.2024.102555.

Kurniawan, M. H., H. Handiyani, T. Nuraini, R. T. S. Hariyati, and S. Sutrisno. 2024. A systematic review of artificial intelligence-powered (AI-powered) chatbot intervention for managing chronic illness. *Annals of Medicine* 56(1). https://doi.org/10.1080/07853890.2024.2302980.

Kwong, J. C. C., G. C. Nickel, S. C. Y. Wang, and J. C. Kvedar. 2024. Integrating artificial intelligence into healthcare systems: More than just the algorithm. *NPJ Digital Medicine* 7(1):52. https://doi.org/10.1038/s41746-024-01066-z.

Lasko, T. A., E. V. Strobl, and W. W. Stead. 2024. Why do probabilistic clinical models fail to transport between sites. *NPJ Digital Medicine* 7:53. https://doi.org/10.1038/s41746-024-01037-4.

Leckenby, E., D. Dawoud, J. Bouvy, and P. Jónsson. 2021. The sandbox approach and its potential for use in health technology assessment: A literature review. *Applied*

Health Economics and Health Policy 19(6):857–869. https://doi.org/10.1007/s40258-021-00665-1.

Ledley, R. S., and L. B. Lusted 1959. Reasoning foundations of medical diagnosis. *Science* 130:9–21. https://doi.org/10.1126/science.130.3366.9.

Lee, J., S. Patel, and A. Taxter. 2023. How to make the electronic health record your friend. *Current Opinion in Pediatrics* 35(5):579–584. https://doi.org/10.1097/MOP.0000000000001261.

Leslie, D., A. Mazumder, A. Peppin, M. K. Wolters, and A. Hagerty. 2021. Does "AI" stand for augmenting inequality in the era of covid-19 healthcare? *BMJ* 372:304. https://doi.org/10.1136/bmj.n304.

Levin, C., E. Naimi, and M. Saban. 2024. Evaluating GenAI systems to combat mental health issues in healthcare workers: An integrative literature review. *International Journal of Medical Informatics* 191:105566. https://doi.org/10.1016/j.ijmedinf.2024.105566.

Levy, D. R., M. Hobensack, K. Cato, D. E. Detmer, K. B. Johnson, J. Williamson, J. Murphy, A. Moy, J. Withall, R. Lee, S. C. Rossetti, and S. T. Rosenbloom. 2022. 25 × 5 Symposium to reduce documentation burden: Report-out and call for action. *Applied Clinical Informatics* 13(2):439–446. https://doi.org/10.1055/s-0042-1746169.

Li, F., N. Ruijs, and Y. Lu. 2022. Ethics & AI: A systematic review on ethical concerns and related strategies for designing with AI in health care. *AI* 4(1):28–53. https://doi.org/10.3390/ai4010003.

Li, J., B. J. Cairns, J. Li, and T. Zhu. 2023. Generating synthetic mixed-type longitudinal electronic health records for artificial intelligent applications. *NPJ Digital Medicine* 6(1):98. https://doi.org/10.1038/s41746-023-00834-7.

Lindsell, C. J., W. W. Stead, and K. B. Johnson. 2020. Action-informed artificial intelligence-matching the algorithm to the problem. *JAMA* 323(21):2141–2142. https://doi.org/10.1001/jama.2020.5035.

Lion, K. C., Y. Lin, and T. Kim. 2024. Artificial intelligence for language translation: The equity is in the details. *JAMA* 332(17):1427–1428. https://doi.org/10.1001/jama.2024.15296.

Liu, J., C. Wang, and S. Liu. 2023. Utility of chatGPT in clinical practice. *Journal of Medical Internet Research* 25:e48568. https://doi.org/10.2196/48568.

Liu, K., and D. Tao. 2022. The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart health care services. *Computers in Human Behavior* 127:107026. https://doi.org/10.1016/j.chb.2021.107026.

Lohr, S. 2005. Health industry under pressure to computerize. *The New York Times*. https://www.nytimes.com/2005/02/19/business/health-industry-under-pressure-to-computerize.html?searchResultPosition=1 (accessed December 9. 2024).

Lopes, L., A. Montero, M. Presiado, and L. Hamel. 2024. Americans' challenges with health care costs. KFF Issue Brief. https://www.kff.org/health-costs/issue-brief/americans-challenges-with-health-care-costs (accessed April 4, 2025).

Lorenz, E. N. 1972. Predictability: Does the flap of a butterfly's wings in Brazil set off a tornado in Texas? Presented at the 139th Annual Meeting of the American Association for the Advancement of Science.

Lucian Leape Institute. 2024. *Patient Safety and Artificial Intelligence: Opportunities and Challenges for Care Delivery.* Institute for Healthcare Improvement. https://www.ihi.org/resources/publications/patient-safety-and-artificial-intelligence-opportunities-and-challenges-care (accessed April 4, 2025).

Lynch, K., M. Kendall, K. Shanks, A. Haque, E. Jones, M. G. Wanis, M. Furukawa, and F. Mostashari. 2014. The health IT regional extension center program: Evolution and lessons for health care transformation. *Health Services Research* 49:421–437. https://doi.org/10.1111/1475-6773.12140.

Malgaroli, M., E. Tseng, T. D. Hull, E. Jennings, T. K. Choudhury, and N. M. Simon. 2023. Association of health care work with anxiety and depression during the COVID-19 pandemic: Structural topic modeling study. *JMIR AI* 2:e47223. https://doi.org/10.2196/47223.

Mandl, K. D., D. Gottlieb, and J. C. Mandel. 2024. Integration of AI in health care requires an interoperable digital data ecosystem. *Nature Medicine* 30:631–634. https://doi.org/10.1038/s41591-023-02783-w.

Mäntymäki, M., M. Minkkinen, T. Birkstedt, and M. Viljanen. 2022. Defining organizational AI governance. *AI and Ethics* 2:603–609. https://doi.org/10.1007/s43681-022-00143-x.

Matheny, M., D. Whicher, and S. Thadaney Israni. 2020. Artificial intelligence in health care: A report from the National Academy of Medicine. *JAMA* 323(6):509–510. https://doi.org/10.1001/jama.2019.21579.

Mazzucato, M., M. Schaake, S. Krier, and J. Entsminger. 2022. *Governing Artificial Intelligence in the Public Interest*. Working Paper Series, UCL Institute for Innovation and Public Purpose.

McCarthy, J., M. L. Minsky, N. Rochester, and C. E. Shannon. 1955. *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html (accessed December 17, 2024).

McCoy, A. B., E. M. Russo, K. B. Johnson, B. Addison, N. Patel, J. P. Wanderer, D. E. Mize, J. G. Jackson, T. J. Reese, S. Littlejohn, L. Patterson, T. French, D. Preston, A. Rosenbury, C. Valdez, S. D. Nelson, C. V. Aher, M. W. Alrifai, J. Andrews, C.

Cobb, S. N. Horst, D. P. Johnson, L. A. Knake, A. A. Lewis, L. Parks, S. K. Parr, P. Patel, B. L. Patterson, C. M. Smith, K. D. Suszter, R. W. Turer, L. J. Wilcox, A. P. Wright, and A. Wright. 2022. Clinician collaboration to improve clinical decision support: The Clickbusters initiative. *Journal of the American Medical Informatics Association* 29(6):1050–1059. https://doi.org/10.1093/jamia/ocac027.

McCradden, M. D., S. Joshi, J. A. Anderson, and A. J. London. 2023. A normative framework for artificial intelligence as a sociotechnical system in healthcare. *Patterns* 4(11):100864. https://doi.org/10.1016/j.patter.2023.100864.

McGenity, C., E. L. Clarke, C. Jennings, G. Matthews, C. Cartlidge, H. Freduah-Agyemang, D. D. Stocken, and D. Treanor. 2024. Artificial intelligence in digital pathology: A systematic review and meta-analysis of diagnostic test accuracy. *NPJ Digital Medicine* 7:114. https://doi.org/10.1038/s41746-024-01106-8.

McGinnis, J. M., H. V. Fineberg, and V. J. Dzau. 2024. Shared commitments for health and health care: A trust framework from the Learning Health System. *NAM Perspectives.* Commentary, Washington, DC. National Academy of Medicine. https://doi.org/10.31478/202412c.

McGlynn, E. A. 2020. Measuring and improving quality in the US: Where are we today? *Journal of the American Board of Family Medicine* 33(Suppl):S28–S35. https://doi.org/10.3122/jabfm.

Mello, M., and N. Guha. 2024. Understanding liability risk from using health care artificial intelligence tools. *New England Journal of Medicine* 271–278.

Melnick, E. R., L. N. Dyrbye, C. A. Sinsky, M. Trockel, C. P. West, L. Nedelec, M. A. Tutty, and T. Shanafelt. 2020. The association between perceived electronic health record usability and professional burnout among US physicians. *Mayo Clinic Proceedings* 95(3):476–487. https://doi.org/10.1016/j.mayocp.2019.09.024.

Mikkelsen, J. G., N. L. Sorensen, C. H. Merrild, M. B. Jensen, and J. L. Thomsen. 2023. Patient perspectives on data sharing regarding implementing and using artificial intelligence in general practice—a qualitative study. *BMC Health Services Research* 23:335. https://doi.org/10.1186%2Fs12913-023-09324-8.

Mills, S., A. Casovan, and V. Shankar. 2023. *A Guide to AI Governance for Business Leaders.* Boston Consulting Group. https://www.bcg.com/publications/2023/a-guide-to-mitigating-ai-risks (accessed April 4, 2025).

Mirsky, Y., and W. Lee. 2021. The creation and detection of deepfakes: A survey. *ACM Computing Surveys* (CSUR) 54(1):7.

MITRE. 2024. *Artificial Intelligence Maturity Model.* https://www.mitre.org/news-insights/fact-sheet/artificial-intelligence-maturity-model (accessed April 4, 2025).

Mortensen, O. 2024. How Many Users Does ChatGPT Have? Statistics & Facts. *SEO News.* https://seo.ai/blog/how-many-users-does-chatgpt-have (accessed July 9, 2024).

Moy, A. J., J. M. Schwartz, R. Chen, S. Sadri, E. Lucas, K. D. Cato, and S. Collins Rossetti. 2021. Measurement of clinical documentation burden among physicians and nurses using electronic health records: A scoping review. *Journal of the American Medical Informatics Association* 28(5):998–1008. https://doi.org/10.1093/jamia/ocaa325.

Mullowney, M. W., K. R. Duncan, S. S. Elsayed, N. Garg, J. J. J. van der Hooft, N. I. Martin, D. Meijer, B. R. Terlouw, F. Biermann, K. Blin, J. Durairaj, M. Gorostiola González, E. J. N. Helfrich, F. Huber, S. Leopold-Messer, K. Rajan, T. de Rond, J. A. van Santen, M. Sorokina, M. J. Balunas, M. A. Beniddir, D. A. van Bergeijk, L. M. Carroll, C. M. Clark, D.-A. Clevert, C. A. Dejong, C. Du, S. Ferrinho, F. Grisoni, A. Hofstetter, W. Jespers, O.V. Kalinina, S. A. Kautsar, H. Kim, T. F. Leao, J. Masschelein, E. R. Rees, R. Reher, D. Reker, P. Schwaller, M. Segler, M. A. Skinnider, A. S. Walker, E. L. Willighagen, B. Zdrazil, N. Ziemert, R. J. M. Goss, P. Guyomard, A. Volkamer, W. H. Gerwick, H. U. Kim, R. Müller, G. P. van Wezel, G. J. P. van Westen, A. K. H. Hirsch, R. G. Linington, S. L. Robinson, and M. H. Medema. 2023. Artificial intelligence for natural product drug discovery. *Nature Reviews* 22:895–916. https://doi.org/10.1038/s41573-023-00774-7.

Mumtaz, H., M. K. Ejaz, M. Tayyab, L. Vohra, S. Sapkota, M. Hasan, and M. Saqib. 2023. APACHE scoring as an indicator of mortality rate in ICU patients: A cohort study. *Annals of Medicine and Surgery (London)* 85(3):416–421. https://doi.org/10.1097/ms9.0000000000000264.

Murdoch, B. 2021. Privacy and artificial intelligence: Challenges for protecting health information in a new era. *BMC Medical Ethics* 22:122. https://doi.org/10.1186/s12910-021-00687-3.

Muro, M., and S. Liu. 2021. *How to Prevent a Winner-Take-Most Outcome for the U.S. AI Economy*. The Brookings Institution. https://www.brookings.edu/articles/how-to-prevent-a-winner-take-most-outcome-for-the-u-s-ai-economy (accessed April 4, 2025).

Nan, J., M. S. Herbert, S. Purpura, A. N. Henneken, D. Ramanathan, and J. Mishra. 2024. Personalized machine learning-based prediction of wellbeing and empathy in healthcare professionals. *Sensors (Basel)* 24(8):2640. https://doi.org/10.3390%2Fs24082640.

National Academy of Medicine (NAM). 2019. *Taking Action Against Clinician Burnout: A Systems Approach to Professional Well-Being.* Washington, DC: The National Academies Press.

NAM. 2022a. *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. Washington, DC: The National Academies Press.

NAM. 2022b. Assessing Meaningful Community Engagement: A Conceptual Model to Advance Health Equity Through Transformed Systems for Health. *Perspectives.* Commentary. https://doi.org/10.31478/202202c.

NAM. 2022c. *National Plan for Health Workforce Well-Being*. Washington, DC: The National Academies Press.

NAM. n.d. *Health Care Artificial Intelligence Code of Conduct*. https://nam.edu/programs/value-science-driven-health-care/health-care-artificial-intelligence-code-of-conduct (accessed July 11, 2024).

National Association for Healthcare Quality (NAHQ). 2022. *Healthcare Quality and Safety Workforce Report: New Imperatives for Quality and Safety Mean New Imperatives for Workforce Development*. https://nahq.org/nahq-intelligence/insights/#report (accessed July 30, 2024).

National Conference of State Legislatures. 2024. *Artificial Intelligence 2023 Legislation*. https://www.ncsl.org/technology-and-communication/artificial-intelligence-2023-legislation (accessed July 20, 2024).

National Institutes of Health (NIH). 2024. *Toward an Ethical Framework for Artificial Intelligence in Biomedical and Behavioral Research: Transparency for Data and Model Reuse*. https://datascience.nih.gov/sites/default/files/NIH-Transparency-Workshop-Report-v7-FINAL-updated-2-12-11-24-508.pdf (accessed December 12, 2024).

National Research Council (NRC). 2009. *Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions.* Washington, DC: The National Academies Press.

National Science Foundation (NSF). n.d. *Artificial Intelligence.* https://new.nsf.gov/focus-areas/artificial-intelligence (accessed April 4, 2025).

NEJM-AI Grand Rounds. 2023. *Examining Open-Source Frameworks and AI in Medicine with Interventional Radiologist Dr. Judy Gichoya*. Narrated by Arjun Manrai and Andrew Beam. https://nejm.ai/ep12 (accessed April 4, 2025).

Nijor, S., G. Rallis, N. Lad, and E. Gokcen. 2022. Patient safety issues from information overload in electronic medical records. *Journal of Patient Safety* 18(6):999–1003. https://doi.org/10.1097%2FPTS.0000000000001002.

Nilsen, P., D. Sundemo, F. Heintz, M. Neher, J. Nygren, P. Svedberg, and L. Petersson. 2024. Towards evidence-based practice 2.0: Leveraging artificial intelligence in health care. *Frontiers in Health Services* 4:1368030. https://doi.org/10.3389/frhs.2024.1368030.

Nobel Prize Outreach AB. 2024. *The Nobel Prize*. Press release. https://www.nobelprize.org/prizes/physics/2024/press-release (accessed December 9, 2024).

NORC. 2021. *Surveys of Trust in the U.S. Health Care System*. https://www.norc.org/content/dam/norc-org/pdfs/20210520_NORC_ABIM_Foundation_Trust%20in%20Healthcare_Part%201.pdf (accessed April 4, 2025).

Nori, H., N. King, S. M. McKinney, D. Carignan, and E. Horvitz. 2023. Capabilities of GPT-4 on medical challenge problems. *arXiv*. https://doi.org/10.48550/arXiv.2303.13375.

Obermeyer, Z., and E. J. Emanuel. 2016. Predicting the future—big data, machine learning, and clinical medicine. *New England Journal of Medicine* 375(13):1216–1219. https://doi.org/10.1056/nejmp1606181.

Obermeyer, Z., B. Powers, C. Vogeli, and S. Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464):447–453. https://doi.org/10.1126/science.aax2342.

O'Brien, M. 2024. *US Ahead in AI Innovation, Easily Surpassing China in Stanford's New Ranking*. AP News, November 21. https://apnews.com/article/ai-us-china-competition-stanford-index-uk-india-c8eb9be0253eb39776c3e38d05f1a329 (accessed April 4, 2025).

Office for Civil Rights (OCR). n.d. *Breach Portal: Notice to the Secretary of HHS breach of unsecured protected health information*. https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf;jsessionid=918B4E4145778C463E32915B79A40F34 (accessed April 4, 2025).

Office of Inspector General, Department of Health and Human Services (OIG HHS). 2022. *Adverse Events in Hospitals: A Quarter of Medicare Patients Experienced Harm in October 2018* (OEI-06-18-00400). https://oig.hhs.gov/oei/reports/OEI-06-18-00400.pdf (accessed April 4, 2025).

Office of the National Coordinator for Health Information Technology (ONC). 2010. Health information technology: Initial set of standards, implementation specifications, and certification criteria for electronic health record technology. Final rule, *Federal Register* 75:44590–44654.

ONC. 2011. Establishment of the permanent certification program for health information technology. Final rule. *Federal Register* 76:1262–1331.

ONC. 2016. ONC health IT certification program: Enhanced oversight and accountability. Final rule. *Federal Register* 81:72404–72471.

ONC. 2024a. Health data, technology, and interoperability: Certification program updates, algorithm transparency, and information sharing. Final rule. *Federal Register* 89:1192–1438.

ONC. 2024b. HTI-1 final rule overview. Presented at the HITAC Meeting. https://www.healthit.gov/sites/default/files/facas/2024-01-18_HTI-1_Final_Rule_Overview_508.pdf (accessed April 4, 2025).

ONC. n.d. *National Trends in Hospital and Physician Adoption of Electronic Health Records.* https://www.healthit.gov/data/quickstats/national-trends-hospital-and-physician-adoption-electronic-health-records (accessed July 23, 2024).

Oikonomou, E. K., and R. Khera. 2024. Designing medical artificial intelligence systems for global use: Focus on interoperability, scalability, and accessibility. *Hellenic Journal of Cardiology* 81:9–17. https://doi.org/10.1016/j.hjc.2024.07.003.

Olawade, D. B., O. Z. Wada, A. Odetayo, A. C. David-Olawade, F. Asaolu, and J. Eberhardt. 2024. Enhancing mental health with artificial intelligence: Current trends and future prospects. *Journal of Medicine, Surgery, and Public Health* 3:100099. https://doi.org/10.1016/j.glmedi.2024.100099.

O'Malley, A., K. Draper, R. Gourevitch, D. Cross, and S. H. Scholle. 2015. Electronic health records and support for primary care teamwork. *Journal of the American Medical Informatics Association* 22(2):426–434. https://doi.org/10.1093/jamia/ocu029.

Organisation for Economic Co-operation and Development (OECD). 2023. Regulatory sandboxes in artificial intelligence. *OECD Digital Economy Papers* 356. https://doi.org/10.1787/8f80a0e6-en.

OECD. 2024. *OECD AI Principles Overview.* https://oecd.ai/en/ai-principles (accessed April 4, 2025).

O'Sullivan, L., R. Crowley, E. McAuliffe, and P. Doran. 2020. Contributory factors to the evolution of the concept and practice of informed consent in clinical research: A narrative review. *Contemporary Clinical Trials Communications* 19:100634. https://doi.org/10.1016/j.conctc.2020.100634.

Overgaard, S. M., M. G. Graham, T. Brereton, M. J. Pencina, J. D. Halamka, D. E. Vidal, and N. J. Economou-Zavlanos. 2023. Implementing quality management systems to close the AI translation gap and facilitate safe, ethical, and effective health AI solutions. *NPJ Digital Medicine* 6:218. https://doi.org/10.1038/s41746-023-00968-8.

Parag, N., R. Govender, and S. B. Ally. 2023. Promoting cultural inclusivity in healthcare artificial intelligence: A framework for ensuring diversity. *Health Management, Policy and Innovation* 8:3.

Paranjape, K., M. Schinkel, R. Nannan Panday, J. Car, and P. Nanayakkara. 2019. Introducing artificial intelligence training in medical education. *JMIR Medical Education* 5(2):e16048. https://doi.org/10.2196/16048.

Parikh, R. B., and L. A. Helmchen. 2022. Paying for artificial intelligence in medicine. *NPJ Digital Medicine* 5:63. https://doi.org/10.1038/s41746-022-00609-6.

Patient-Centered Outcomes Research Institute (PCORI). n.d. *Engagement in Research: PCORI's Foundational Expectations for Partnerships*. https://www.pcori.org/engagement/engagement-resources/engagement-research-pcoris-foundational-expectations-partnerships (accessed April 4, 2025).

Patterson, A. M., M. O'Boyle, G. E. VanNoy, and K. A. Dies. 2023. Emerging roles and opportunities for rare disease patient advocacy groups. *Therapeutic Advances in Rare Disease* 4. https://doi.org/10.1177%2F26330040231164425.

Pearce, R., Y. Li, G. S. Omenn, and Y. Zhang. 2022. Fast and accurate Ab Initio Protein structure prediction using deep learning potentials. *PLoS Computational Biology* 18(9):e1010539. https://doi.org/10.1371/journal.pcbi.1010539.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Francisco, CA: Morgan Kaufmann.

Peccoralo, L., C. Kaplan, R. Pietrzak, D. Charney, and J. Ripp. 2021. The impact of time spent on the electronic health record after work and of clerical work on burnout among clinical faculty. *Journal of the American Medical Informatics Association* 28(5):938–947. https://doi.org/10.1093/jamia/ocaa349.

Perković, G., A. Drobnjak, and I. Botički. 2024. Hallucinations in LLMs: Understanding and addressing challenges. *47th MIPRO ICT and Electronics Convention (MIPRO).* Opatija, Croatia. IEEE, pp. 2084–2088. https://doi.org/10.1109/MIPRO60963.2024.10569238.

Perni, S., L. S. Lehmann, and D. S. Bitterman. 2023. Patients should be informed when AI systems are used in clinical trials. *Nature Medicine* 29(8):1890–1891. https://doi.org/10.1038/s41591-023-02367-8.

Petersson, L., I. Larsson, J. Nguyen, P. Nilsen, M. Neher, J. E. Reed, D. Tyskbo, and P. Svedberg. 2022. Challenges to implementing artificial intelligence in health care: A qualitative interview study with health care leaders in Sweden. *BMC Health Services Research* 22(1):850. https://doi.org/10.1186/s12913-022-08215-8.

Phillips, J., and J. D. Klein. 2023. Change management: From theory to practice. *TechTrends* 67(1):189–197. https://doi.org/10.1007/s11528-022-00775-0.

Pittman, P., C. Chen, C. Erikson, Q. Luo, A. Vichare, S. Batra, and G. Burke. 2021. Health workforce for health equity. *Medical Care* 59:405–408. https://doi.org/10.1097%2FMLR.0000000000001609.

Placani, A. 2024. Anthropomorphism in AI: Hype and fallacy. *AI Ethics* 4:691–698. https://doi.org/10.1007/s43681-024-00419-4.

Pouyan, E. 2024. Challenges and strategies for wide-scale artificial intelligence (AI) deployment in healthcare practices: A perspective for healthcare organizations. *Artificial Intelligence in Medicine* 151:102861. https://doi.org/10.1016/j.artmed.2024.102861.

Pylypchuk, Y., and C. Johnson. 2022. New EHR certification requirements and their association with duplicate tests and images. *Journal of the American Medical Informatics Association* 29(8):1391–1399. http://dx.doi.org/10.1093/jamia/ocac076.

Qiang, J., D. Wu, H. Du, H. Zhu, S. Chen, and H. Pan. 2022. Review on facial-recognition-based applications in disease diagnosis. *Journal of Personalized Medicine* 9(7):273. http://dx.doi.org/10.3390/bioengineering9070273.

Quinn, T. P., M. Senadeera, S. Jacobs, S. Coghlan, and V. Le. 2021. Trust and medical AI: The challenges we face and the expertise needed to overcome them. *Journal of the American Medical Informatics Association* 28(4):890–894. https://doi.org/10.1093/jamia/ocaa268.

Rajaguru, V., W. Han, T. H. Kim, I. Shin, and S. G. Lee. 2022. LACE index to predict the high risk of 30-day readmission: A systematic review and meta-analysis. *Journal of Personalized Medicine* 12(4):545. https://doi.org/10.3390/jpm12040545.

Raji, I. D., A. Smart, R. N. White, M. Mitchell, T. Gebru, B. Hutchinson, J. Smith-Loud, D. Theron, and P. Barnes. 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency.* New York: Association for Computing Machinery, pp. 33–44. https://doi.org/10.1145/3351095.3372873.

Rajkomar, A., J. Dean, and I. Kohane. 2019. Machine learning in medicine. *New England Journal of Medicine* 380(14):1347–1358. https://doi.org/10.1056/nejmra1814259.

Ratwani, R. M., D. Classen, and C. Longhurst. 2024. The compelling need for shared responsibility of AI oversight: Lessons from health IT certification. *JAMA* 332(10):787–788. https://doi.org/10.1001/jama.2024.12630.

Reddy, S. 2024. Generative AI in healthcare: An implementation science informed translational path on application, integration and governance. *Implementation Science* 19:27. https://doi.org/10.1186/s13012-024-01357-9.

Reddy, S. n.d. The impact of AI on the health care workforce: Balancing opportunities and challenges. *Health Information Management Systems Society*. https://www.himss.org/resources/impact-ai-health care-workforce-balancing-opportunities-and-challenges (accessed July 14, 2024).

Reddy, S., W. Rogers, V. Makinen, E. Coiera, P. Brown, M. Wenzel, E. Weicken, S. Ansari, P. Mathur, A. Casey, and B. Kelly. 2021. Evaluation framework to guide implementation of AI systems into healthcare settings. *BMJ Health Care Informatics* 28(1):e100444. https://doi.org/10.1136/bmjhci-2021-100444.

Reed, M., J. Huang, I. Graetz, R. Brand, J. Hsu, B. Fireman, and M. Jaffe. 2012. Outpatient electronic health records and the clinical care and outcomes of patients with diabetes mellitus. *Annals of Internal Medicine* 157(7):482–489. https://doi.org/10.7326/0003-4819-157-7-201210020-00004.

Reuters. 2023. *China's Xi Calls for Accelerated BRICS Expansion*. https://www.reuters.com/world/chinas-xi-calls-accelerated-brics-expansion-2023-08-23 (accessed April 4, 2025).

Riddell, M. C., K. G. Sandford, A. O. Johnson, C. Steltenkamp, and K. A. Pearce. 2014. Achieving meaningful use of electronic health records (EHRs) in primary care: Proposed critical processes from the Kentucky Ambulatory Network (KAN). *Journal of the American Board of Family Medicine* 27(6):772–779. https://doi.org/10.3122/jabfm.2014.06.140030.

Roberts, H., E. Hine, M. Taddeo, and L. Floridi. 2024. Global AI governance—barriers and pathways forward. *International Affairs* 1275–1286. https://doi.org/10.1093/ia/iiae073.

Robinson, K. E., and J. A. Kersey. 2018. Novel electronic health record (EHR) education intervention in large health care organization improves quality, efficiency, time, and impact on burnout. *Medicine (Baltimore)* 97(38):e12319. https://doi.org/10.1097/MD.0000000000012319.

Rumsfeld, D. 2011. *Known and Unknown: A Memoir*. Penguin.

Saeed, W., and C. Omlin. 2023. Explainable AI (XAI): A systematic meta survey of current challenges and future opportunities. *Knowledge-Based Systems* 263:110273. https://doi.org/10.1016/j.knosys.2023.110273.

Sahini, N., G. Stein, R. Zemmel, and D. M. Cutler. 2023. *The potential impact of artificial intelligence on healthcare spending.* Working Paper 30857. National Bureau of Economic Research. https://doi.org/10.3386/w30857.

Sallam, M. 2023. ChatGPT utility in health care education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Health Care (Basel)* 11(6):887. https://doi.org/10.3390/health care11060887.

Salwei, M., and P. Carayon. 2022. A sociotechnical systems framework for the application of artificial intelligence in health care delivery. *Journal of Cognitive Engineering and Decision Making* 16(4):194–206. https://doi.org/10.1177/15553434221097357.

Sandalow, M. 2024. *First into the Breach: ONC Final Rule Addressing AI Transparency in Health Care*. Bipartisan Policy Center. https://bipartisanpolicy.org/blog/first-into-the-breach-onc-final-rule-addressing-ai-transparency-in-health-care (accessed April 4, 2025).

Sanders, H. 2024. 260 McNuggets? McDonald's ends AI drive-through tests amid errors. *The New York Times*. https://www.nytimes.com/2024/06/21/business/

mcdonalds-ai-drive-thru-white-castle.html?smid=url-share (accessed April 4, 2025).

Sanderson, S. C., K. B. Brothers, N. D. Mercaldo, E. W. Clayton, A. H. Matheny Antommaria, S. A. Aufox, M. H. Brilliant, D. Campos, D. S. Carrell, J. Connolly, P. Conway, S. M. Fullerton, N. A. Garrison, C. R. Horowitz, G. P. Jarvik, D. Kaufman, T. E. Kitchner, R. Li, E. J. Ludman, C. A. McCarty, and I. A. Holm. 2017. Public attitudes toward consent and data sharing in biobank research: A large multi-site experimental survey in the US. *American Journal of Human Genetics* 100(3):414–427. https://doi.org/10.1016/j.ajhg.2017.01.021.

Sandmann, S., S. Riepenhausen, L. Plagwitz, and J. Varghese. 2024. Systematic analysis of ChatGPT, Google search and Llama 2 for clinical decision support tasks. *Nature Communications* 15:2050. https://doi.org/10.1038/s41467-024-46411-8.

Sanyal, S. 2021. How much does artificial intelligence cost in 2021? *Analytics Insight*. https://www.analyticsinsight.net/artificial-intelligence/how-much-does-artificial-intelligence-cost-in-2021 (accessed April 4, 2025).

Sarasohn-Kahn, J. 2021. *A Nutrition Label for Health IT*. HEALTHPOPULI blog. https://www.healthpopuli.com/2021/09/08/a-nutrition-label-for-health-it (accessed April 4, 2025).

Saria, S. 2022. Not all AI is created equal: Strategies for safe and effective adoption. *NEJM Catalyst Innovations in Care Delivery* 3(2). https://doi.org/10.1056/CAT.22.0075.

Sartori, L., and A. Theodorou. 2022. A sociotechnical perspective for the future of AI: Narratives, inequalities, and human control. *Ethics and Information Technology* 24:4. https://doi.org/10.1007/s10676-022-09624-3.

Sauerbrei, A., A. Kerasidou, F. Lucivero, and N. Hallowell. 2023. The impact of artificial intelligence on the person-centred, doctor-patient relationship: Some problems and solutions. *BMC Medical Informatics and Decision Making* 23:73. https://doi.org/10.1186/s12911-023-02162-y.

Schmitt, L. 2022. Mapping global AI governance. *AI and Ethics* 2:303–314. https://doi.org/10.1007/s43681-021-00083-y.

Scott, I. A., S. M. Carter, and E. Coiera. 2021. Exploring stakeholder attitudes towards AI in clinical practice. *BMJ Health & Care Informatics.* 28:e100450. https://doi.org/10.1136/bmjhci-2021-100450.

Sellen, A., and E. Horvitz. 2024. The rise of the AI co-pilot: Lessons for design from aviation and beyond. *Communications of the ACM* 67(7):18–23. https://doi.org/10.1145/3637865.

Semprini, J. 2023. Examining racial disparities in unemployment among health care workers before, during, and after the COVID-19 pandemic.

Journal of Patient-Centered Research and Reviews 10(3):136–141. https://doi.org/10.17294%2F2330-0698.2021.

Sendak, M. P., M. Gao, N. Brajer, and S. Balu. 2020. Presenting machine learning model information to clinical end users with model facts labels. *NPJ Digital Medicine* 3:41. https://doi.org/10.1038/s41746-020-0253-3.

Sendak, M. P., S. Balu, and W. N. Price ll. 2023. Enabling collaborative governance of medical AI. *Nature Machine Intelligence* 5:821–823. https://doi.org/10.1038/s42256-023-00699-1.

Sepasspour, R. 2023. A reality check and a way forward for the global governance of artificial intelligence. *Bulletin of the Atomic Scientists* 79(5):304–315. https://doi.org/10.1080/00963402.2023.2245249.

Shah, N. H., J. D. Halamka, S. Saria, M. Percina, T. Tazbaz, M. Tripathi, A. Callahan, H. Hildahl, and B. Anderson. 2024. A nationwide network of health AI assurance laboratories. *JAMA* 331(3):245–249. https://doi.org/10.1001/jama.2023.26930.

Shanafelt, T. D. 2021. Physician well-being 2.0: Where are we and where are we going? *Mayo Clinic Proceedings* 96(10):2682–2693. https://doi.org/10.1016/j.mayocp.2021.06.005.

Shearer, C. 2000. The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing* 5:13–22.

Shickel, B., P. J. Tighe, A. Bihorac, and P. Rashidi. 2018. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE Journal of Biomedical and Health Informatics* 22(5):1589–1604. https://doi.org/10.1109/jbhi.2017.2767063.

Siegel, R., K. Gordon, and L. Dynan. 2021. Behavioral economics: A primer and applications to the UN Sustainable Development Goal of good health and well-being. *EMBO Reports* 4(2):16. https://doi.org/10.3390/reports4020016.

Silcox, C., E. Zimlichmann, K. Huber, N. Rowen, R. Saunders, M. McClellan, C. N. Kahn III, C. A. Salzberg, and D. W. Bates. 2024. The potential for artificial intelligence to transform healthcare: Perspectives from international health leaders. *NPJ Digital Medicine* 7:88. https://doi.org/10.1038/s41746-024-01097-6.

Siontis, G. C. M., R. Sweda, P. A. Noseworthy, P. A. Friedman, K. C. Siontis, and C. J. Patel. 2021. Development and validation pathways of artificial intelligence tools evaluated in randomised clinical trials. *BMJ Health Care Informatics* 28(1):e100466. https://doi.org/10.1136%2Fbmjhci-2021-100466.

Sjoding, M. W., R. P. Dickson, T. Iwashyna, S. E. Gay, and T. S. Valley. 2020. Racial bias in pulse oximetry measurement. *New England Journal of Medicine* 383:2477–2478. https://doi.org/10.1056/nejmc2029240.

Skovgaard, L. L., S. Wadmann, and K. Hoeyer. 2019. A review of attitudes towards the reuse of health data among people in the European Union: The primacy of purpose and the common good. *Health Policy* 123(6):564–571. https://doi.org/10.1016%2Fj.healthpol.2019.03.012.

Steerling, E., E. Siira, P. Nilsen, P. Svedberg, and J. Nygren. 2023. Implementing AI in health care—The relevance of trust: A scoping review. *Frontiers in Health Services* 3:1211150. https://doi.org/10.3389/frhs.2023.1211150.

Steimetz, E., J. Minkowitz, E. C. Gabutan, J. Ngichabe, H. Attia, M. Hershkop, F. Ozay, M. G. Hanna, and R. Gupta. 2024. Use of artificial intelligence chatbots in interpretation of pathology reports. *JAMA Network Open* 7(5):e2412767. https://doi.org/10.1001/jamanetworkopen.2024.12767.

Stern, A. D., A. Goldfarb, T. Minssen, and W. N. Price. 2022. AI insurance: How liability insurance can drive the responsible adoption of artificial intelligence in health care. *NEJM Catalyst* 3(4). http://dx.doi.org/10.1056/CAT.21.0242.

Subbaswamy, A., and S. Saria. 2020. From development to deployment: Dataset shift, causality, and shift-stable models in health AI. *Biostatistics* 21(1):345–352. https://doi.org/10.1093/biostatistics/kxz041.

Suresh, H., and J. V. Guttag. 2021. A framework for understanding unintended consequences of machine learning. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. New York: Association for Computing Machinery. https://doi.org/10.48550/arXiv.1901.10002.

Sutton, R., D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker. 2020. An overview of clinical decision support systems: Benefits, risks, and strategies for success. *NPJ Digital Medicine* 3:17. https://doi.org/10.1038%2Fs41746-020-0221-y.

Tabassi, E. 2023. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1.

Taleb, N. N. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.

Tallberg, J., E. Erman, M. Furendal, J. Geith, M. Klamberg, and M. Lundgren. 2023. The global governance of artificial intelligence: Next steps for empirical and normative research. *International Studies Review* 25(3):viad040. https://doi.org/10.1093/isr/viad040.

Tang, H., and J. H. K. Ng. 2006. Googling for a diagnosis—use of Google as a diagnostic aid: Internet based study. *BMJ* 333(7579):1143–1145. https://doi.org/10.1136/bmj.39003.640567.AE.

Tao, W., A. N. Concepcion, M. Vianen, A. C. A. Marijnissen, F. P. G. J. Lafeber, T. R. D. J. Radstake, and A. Pandit. 2020. Multiomics and machine learning accurately predict clinical response to adalimumab and etanercept therapy in patients with rheumatoid arthritis. *Arthritis & Rheumatology* 73(2):212–222. https://doi.org/10.1002/art.41516.

Temple, S. W. P., and C. G. Rowbottom. 2024. Gross failure rates and failure modes for a commercial AI-based auto-segmentation algorithm in head and neck cancer patients. *Applied Clinical Medical Physics* 25(6):e14273. https://doi.org/10.1002/acm2.14273.

The Joint Commission. n.d. *Responsible Use of Health Data Certification*. https://www.jointcommission.org/what-we-offer/certification/certifications-by-setting/hospital-certifications/responsible-use-of-health-data-certification (accessed April 4, 2025).

The Light Collective. 2024. *Collective AI Rights for Patients*. https://lightcollective.org/wp-content/uploads/2024/06/Collective-AI-Rights-For-Patients-v_2.0.pdf (accessed April 4, 2025). Thompson, C., T. Margo, F. B. Yu, and C. U. Lehmann. 2022. Developing a pediatric EHR testing & certification program in the USA: Initial phase. In *MEDINFO 2021: One World, One Health—Global Partnership for Digital Innovation, Vol. 290.* IOS Press, pp. 1020–1021. https://doi.org/10.3233/SHTI220247.

Tierney, A. A., G. Gayre, B. Hoberman, B. Mattern, M. Ballesca, P. Kipnis, V. Liu, and K. Lee. 2024. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. *NEJM Catalyst Innovations in Care Delivery* 5(3). https://doi.org/10.1056/CAT.23.0404.

Tighe, P., B. M. Gale, and S. E. Mossburg. 2024. Artificial intelligence and patient safety: Promise and challenges. *PSNet*. Agency for Healthcare Research and Quality. https://psnet.ahrq.gov/perspective/artificial-intelligence-and-patient-safety-promise-and-challenges (accessed April 4, 2025).

Topol, E. J. 2019. High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine* 25:44–56. https://doi.org/10.1038/s41591-018-0300-7.

Tornero-Costa, R., A. Martinez-Millana, N. Azzopardi-Muscat, L. Lazeri, V. Traver, and D. Novillo-Ortiz. 2023. Methodological and quality flaws in the use of artificial intelligence in mental health research: Systematic review. *JMIR Mental Health* 10:e42045. https://doi.org/10.2196/42045.

Tran, B., A. Lenhart, R. Ross, and D. A. Dorr. 2019. Burnout and EHR use among academic primary care physicians with varied clinical workloads. *AMIA Joint Summits on Translational Science Proceedings* 2019:136–144.

Trout, K. E., L. W. Chen, F. A. Wilson, H. J. Tak, and D. Palm. 2022. The impact of electronic health records and meaningful use on inpatient quality. *Journal for Healthcare Quality* 44(2):e15–e23. https://doi.org/10.1097/JHQ.0000000000000314.

Tseng, P., R. Kaplan, B. Richman, M. Shah, and K. Schulman. 2018. Administrative costs associated with physician billing and insurance-related activities at an academic health care system. *JAMA* 319(7):691–697. https://doi.org/10.1001/jama.2017.19148.

21st Century Cures Act. 2016. Public Law 114-255. https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf (accessed December 11, 2024).

U.S. Food and Drug Administration (FDA). 2021. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan.

FDA. 2022. *Artificial Intelligence and Machine Learning in Software as a Medical Device.* https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device (accessed April 4, 2025).

FDA. 2024. *Artificial Intelligence & Medical Products: How CBER, CDER, CDRH, and OCP Are Working Together*. https://www.fda.gov/media/177030/download?attachment (accessed July 3, 2024).

FDA. n.d. *Using Intelligence & Machine Learning in the Development of Drug and Biological Products: Discussion Paper and Request for Feedback*. https://www.fda.gov/media/167973/download?attachment;%C2%A0https://www.fda.gov/media/165743/download?attachment;%20and%C2% (accessed July 3, 2024).

U.S. General Services Administration. n.d. *AI Guide for Government: A Living and Evolving Guide to the Application of Artificial Intelligence for the US Federal Government*. https://coe.gsa.gov/coe/ai-guide-for-government/understanding-managing-ai-lifecycle (accessed April 4, 2025).

Van den Bruel, A., T. Haj-Hassan, M. Thompson, F. Buntinx, and D. Mant. 2010. Diagnostic value of clinical features at presentation to identify serious infection in children in developed countries: A systematic review. *Lancet* 375(9717):834–845. https://doi.org/10.1016/s0140-6736(09)62000-6.

Van den Bruel, A., M. Thompson, F. Buntinx, and D. Mant. 2012. Clinicians' gut feeling about serious infections in children: Observational study. *BMJ* 345:e6144. https://doi.org/10.1136/bmj.e6144.

Van der Heijden, A. A., M. D. Abramoff, F. Verbraak, M. V. van Hecke, A. Liem, and G. Nijpels. 2018. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmology* 96(1):63–68. https://doi.org/10.1111/aos.13613.

van Genderen, M. E., D. van de Sande, L. Hooft, A. A. Reis, A. D. Cornet, J. H. F. Oosterhoff, B. J. P. van der Ster, J. Huiskens, R. Townsend, J. van Bommel, D. Gommers, and J. van den Hoven. 2024. Charting a new course in healthcare: Early-stage AI algorithm registration to enhance trust and transparency. *NPJ Digital Medicine* 7(1):119. https://doi.org/10.1038%2Fs41746-024-01104-w.

Vanstone, M., C. Canfield, C. Evans, M. Leslie, M. A. Levasseur, M. MacNeil, M. Pahwa, J. Panday, P. Rowland, S. Taneja, L. Tripp, J. You, and J. Abelson. 2023. Towards conceptualizing patients as partners in health systems: A systematic review and descriptive synthesis. *Health Research Policy and Systems* 21(1):12. https://doi.org/10.1186/s12961-022-00954-8.

Vasey, B., M. Nagendran, B. Campbell, D. A. Clifton, Gary. S. Collins, S. Denaxas, A. K. Denniston, L. Faes, B. Geerts, M. Ibrahim, X. Liu, B. A. Mateen, P. Mathur, M. D. McCradden, L. Morgan, J. Ordish, C. Rogers, S. Saria, D. S. W. Ting, P. Watkinson, W. Weber, P. Wheatstone, and P. McCulloch. 2022. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ* 377:e070904. https://doi.org/10.1136/bmj-2022-070904.

Vijayakumar, A. 2023. Potential impact of artificial intelligence on the emerging world order. *F1000Research* 11:1186. https://doi.org/10.12688/f1000research.124906.2.

Vollmer, S., B. A. Mateen, G. Bohner, F. J. Király, R. Ghani, P. Jonsson, S. Cumbers, A. Jonas, K. S. L. McAllister, P. Myles, D. Granger, M. Birse, R. Branson, K. G. M. Moons, G. S. Collins, J. P. A. Ioannidis, C. Holmes, and H. Hemingway. 2020. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ* 368:l6927. https://doi.org/10.1136/bmj.l6927.

Wallace, M. B., P. Sharma, P. Bhandari, J. East, G. Antonelli, R. Lorenzetti, M. Vieth, I. Speranza, M. Spadaccini, M. Desai, F. J. Lukens, G. Babameto, D. Batista, D. Singh, W. Palmer, F. Ramirez, R. Palmer, T. Lunsford, K. Ruff, E. Bird-Liebermann, V. Ciofoaia, S. Arndtz, D. Cangemi, K. Puddick, G. Derfus, A. S. Johal, M. Barawi, L. Longo, L. Moro, A. Repici, and C. Hassan. 2022. Impact of artificial intelligence on miss rate of colorectal neoplasia. *Gastroenterology* 163(1):295–304. https://doi.org/10.1053/j.gastro.2022.03.007.

Washington, V., K. DeSalvo, F. Mostashari, and D. Blumenthal. 2017. The HITECH era and the path forward. *New England Journal of Medicine* 377:904–906. https://doi.org/10.1056/NEJMp1703370.

Wiens, J., and E. S. Shenoy. 2018. Machine learning for healthcare: On the verge of a major shift in healthcare epidemiology. *Clinical Infectious Diseases* 66(1):149–153. https://doi.org/10.1093/cid/cix731.

Williams, J. 2024. Battle of the bots: As payers use AI to drive denials higher, providers fight back. *Healthcare Financial Management Association*. https://www.hfma.org/revenue-cycle/denials-management/health-systems-start-to-fight-back-against-ai-powered-robots-driving-denial-rates-higher (accessed March 7, 2025).

Wong, A., E. Otles, J. P. Donnelly, A. Krumm, J. McCullough, O. DeTroyer-Cooley, J. Pestrue, M. Phillips, J. Konye, C. Penoza, M.Ghous, and K. Singh. 2021. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Internal Medicine* 181(8):1065–1070. https://doi.org/10.1001/jamainternmed.2021.2626.

World Economic Forum. 2023. *It's Time We Embrace an Agile Approach to Regulating AI*. https://www.weforum.org/agenda/2023/11/its-time-we-embrace-an-agile-approach-to-regulating-ai (accessed April 4, 2025).

World Economic Forum. 2024. *AI Governance Alliance Briefing Paper Series*. https://www.weforum.org/publications/ai-governance-alliance-briefing-paper-series (accessed April 4, 2025).

World Health Organization (WHO). 2016. *Global Strategy on Human Resources for Health: Workforce 2030*. https://iris.who.int/bitstream/handle/10665/250368/9789241511131-eng.pdf (accessed April 4, 2025).

WHO. 2021. *Generating Evidence for Artificial Intelligence-Based Medical Devices: A Framework for Training, Validation and Evaluation*. https://iris.who.int/bitstream/handle/10665/349093/9789240038462-eng.pdf?sequence=1 (accessed April 4, 2025).

Wu, D., S. Chen, Y. Zhang, H. Zhang, Q. Wang, J. Li, Y. Fu, S. Wang, H. Yang, H. Du, H. Zhu, H. Pan, and Z. Shen. 2021. Facial recognition intensity in disease diagnosis using automatic facial recognition. *Journal of Personalized Medicine* 11(11):1172. http://dx.doi.org/10.3390/jpm11111172.

Wu, K., E. Wu, B. Theodorou, W. Liang, C. Mack, L. Glass, J. Sun, and J. Zou. 2023. Characterizing the clinical adoption of medical AI devices through U.S. insurance claims. *NEJM AI* 1(1). https://doi.org/10.1056/AIoa2300030.

Wyatt, R., M. Laderman, L. Botwinick, K. Mate, and J. Whittington. 2016. *Achieving Health Equity: A Guide for Health Care Organizations*. IHI White Paper. Cambridge, MA: Institute for Healthcare Improvement.

Xing, Y., Y. Sun, H. Li, M. Tang, W. Huang, K. Zhang, D. Zhang, D. Zhang, and Q. Ma. 2018. CHA(2)DS(2)-VASc score as a predictor of long-term cardiac outcomes in elderly patients with or without atrial fibrillation. *Clinical Interventions in Aging* 13:497–504. https://doi.org/10.2147/CIA.S147916.

Yao, X., D. R. Rushlow, J. W. Inselman, R. G. McCoy, T. D. Thacher, E. M. Behnken, M. E. Bernard, S. L. Rosas, A. Akfaly, A. Misra, P. E. Molling, J. S. Krien, R. M. Foss, B. A. Barry, K. C. Siontis, S. Kapa, P. A. Pellikka, F. Lopez-Jimenez,

Z. I. Attia, N. D. Shah, P. A. Friedman, and P. A. Noseworthy. 2021. Artificial intelligence–enabled electrocardiograms for identification of patients with low ejection fraction: A pragmatic, randomized clinical trial. *Nature Medicine* 27:815–819. https://doi.org/10.1038/s41591-021-01335-4.

YOLE Group. 2024. *AWS Steps Up AI Competition*. https://www.yolegroup.com/strategy-insights/aws-steps-up-ai-competition (accessed April 4, 2025).

Yordanova, M. Z. 2024. The applications of artificial intelligence in radiology: Opportunities and challenges. *European Journal of Medical and Health Sciences* 6(2):11–14. https://doi.org/10.24018/ejmed.2024.6.2.2085.

Zack, T., E. Lehman, M. Suzgun, J. A. Rodriguez, L. A. Celi, J. Gichoya, D. Jurafsky, P. Szolovits, D. W. Bates, R. E. Abdulnour, A. J. Butte, and E. Alsentzer. 2024. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: A model evaluation study. *Lancet Digital Health* 6(1):E12–E22. https://doi.org/10.1016/s2589-7500(23)00225-x.

Zhu, C., P. K. Attaluri, P. J. Wirth, E. C. Shaffrey, J. B. Friedrich, and V. K. Rao. 2024. Current applications of artificial intelligence in billing practices and clinical plastic surgery. *Plastic and Reconstructive Surgery—Global Open* 12(7):e5939. https://doi.org/10.1097/GOX.0000000000005939.

Zink, A., M. E. Chernew, and H. T. Neprash. 2024. How should Medicare pay for artificial intelligence? *JAMA Internal Medicine* 184(8):863–864. https://doi.org/10.1001/jamainternmed.2024.1648.

Zong, J., and J. N. Matias. 2022. Bartleby: Procedural and substantive ethics in the design of research ethics systems. *Social Media + Society* 8(1). https://doi.org/10.1177/20563051221077021.

AUTHOR INFORMATION

Andrew Bindman is the executive vice president and the chief medical officer for Kaiser Permanente. He is responsible for driving superior quality and equitable health outcomes through the integration of quality innovation, care delivery, data analytics, and research in collaboration with the Permanente Medical Groups. He is also Kaiser Permanente's executive sponsor for the Kaiser Permanente Bernard J. Tyson School of Medicine. Prior to joining Kaiser Permanente, Dr. Bindman spent more than 30 years on the faculty at the University of California, San Francisco (UCSF), where he practiced and taught internal medicine while conducting research on health access and outcomes that resulted in more than 200 published scientific articles. He also served as the director of the Agency for Healthcare Research and Quality in 2016–2017. Dr. Bindman is a graduate of Harvard College and the Icahn School of Medicine at Mount Sinai. A board-certified general internist, he completed his residency in internal medicine at UCSF and was a Robert Wood Johnson Foundation Clinical Scholar at Stanford University. He is an elected member of the National Academy of Medicine and the Association of American Physicians. (**Conflict and Interest Disclosures:** None)

Grace Cordovano is nationally recognized as a leading patient voice in realms of health equity, techquity, and patients' rights and governance of artificial intelligence (AI). Upon recognizing significant unmet needs and challenges in patients' experiences throughout their cancer diagnosis, Dr. Cordovano founded Enlightening Results in 2010. Dr. Cordovano is dedicated to fostering personalized patient advocacy services, specializing in oncology, rare, and catastrophic cases. She has been likened to the "House" of patient navigation, internationally recognized for "navigating the unnavigable." She strategically guides patients and their care partners through survivorship or end-of-life care planning with empathy, ensuring individuals are armed with the most pertinent, medically credible, easy-to-understand information, tools, and technology to make informed decisions about their care. Dr. Cordovano is a champion for palliative care, digital health, clinical trials, and harnessing the power of AI in navigating life with a cancer diagnosis. She amplifies that patient care strategies must give the competitive advantage to the patient, not their diagnosis. Dr. Cordovano completed her master's degree and Ph.D. in biochemistry at the Albert Einstein College of Medicine. She is a board-certified patient advocate via the international credentialing of the Patient Advocate Certification Board. (**Conflict and Interest Disclosures:** None)

Jodi Daniel is a founder and the managing director at Crowell Health Solutions and a partner at Crowell & Moring where she leads the firm's Digital Health Practice. She provides strategic, legal, and policy advice to health care providers and health plans that are bringing innovation into practice, and to health technology clients navigating the dynamic health regulatory environment. Daniel is a leader in digital health and health data policy. She was the founding director of the Office of Policy in the Office of the National Coordinator for Health IT (ONC) at the Department of Health and Human Services (HHS), where she led the agency's federal advisory committees and established national health information technology (IT) policy on privacy, security, consumer e-health, telehealth, and safety and oversight. Daniel developed ONC's regulatory capacity, led the adoption of health IT standards and certification regulations, and advised the Centers for Medicare & Medicaid Services on health IT incentive programs. Prior to her ONC role, Daniel was a key drafter of the original Health Insurance Portability and Accountability Act Privacy and Enforcement Rules and served as HHS's first senior counsel for health IT. Daniel received a B.A. in economics from Tufts University, an M.P.H. from the Johns Hopkins Bloomberg School of Public Health, and a J.D. from the Georgetown University Law Center. She is currently an assistant professor adjunct at Yale Medical School. (**Conflict and Interest Disclosures:** None)

Wyatt Decker is UnitedHealth Group's executive vice president and chief physician. In this role, Dr. Decker serves as the company's lead ambassador working across the enterprise and externally with key stakeholders to further enable and advance accountable models of care. Dr. Decker previously served as the chief executive officer of Optum Health, UnitedHealth Group's national integrated care delivery platform. During his nearly 5-year tenure, he played a vital role in building out and accelerating Optum Health's value-based care delivery capabilities and helping more than 100,000 employed and contracted physicians achieve lower costs and deliver better outcomes for more than 100 million people, including the nearly 4 million people Optum now serves in fully accountable, value-based arrangements. Under his leadership, Optum Health established national platforms for care delivery, home and community care, behavioral care, benefits, chronic disease management solutions, and the Center for Advanced Clinical Solutions, applying cutting-edge technologies to solve some of health care's toughest problems. Dr. Decker holds an M.D. from the Mayo Clinic Alix School of Medicine, an M.B.A. from the Kellogg School of Management at Northwestern University, and a B.S. from the University of California, Santa Cruz. He has published numerous scientific articles and is recognized among the nation's top 100 health care leaders by Modern Healthcare. (**Conflict and Interest Disclosures:** None)

Peter J. Embí is an internationally recognized researcher, educator, and leader in the field of biomedical informatics, with particular emphasis in clinical and translational research informatics and the effective and ethical application of artificial intelligence in health care. Dr. Embí serves as a professor and the chair of the Department of Biomedical Informatics, a professor of medicine, the co-director of the ADVANCE AI Center, and the senior vice-president for research and innovation at the Vanderbilt University Medical Center (VUMC). Prior to joining VUMC, he served as the president and the chief executive officer of the Regenstrief Institute, a professor and the associate dean for informatics and health services research at the Indiana University (IU) School of Medicine, the associate director at Indiana Clinical and Translational Sciences Institute, the vice president for Learning Health Systems at IU Health, and on the faculty of the University of Cincinnati (UC), where he was the founding director of the UC Center for Health Informatics. Among his other leadership roles, he is the president of the American College of Medical Informatics (ACMI), was the past president and the chair of the Board of Directors of the American Medical Informatics Association, served on the Board of Scientific Counselors to the National Library of Medicine and on the National Advisory Council for the Agency for Healthcare Research and Quality. He is a fellow of the American College of Physicians, a fellow of the ACMI, a fellow of the International Academy of Health Sciences Informatics, and an elected member of the National Academy of Medicine. (**Conflict and Interest Disclosures:** None)

Gianrico Farrugia is the president and the chief executive officer of the Mayo Clinic. With a distinct focus on advancing patient-centered health care transformation for the health care sector, Dr. Farrugia is a pioneering voice in moving health care from a linear pipeline operating model to a platform-based model. Under his leadership, the Mayo Clinic launched and scaled Mayo Clinic

Platform, which brings together data partners, solution developers, and health care providers to transform care through insights derived from the world's most comprehensive repository of longitudinal de-identified clinical data across four continents. To further enable transformation, Dr. Farrugia and his leadership team are re-architecting Mayo Clinic's physical infrastructure to support the future of care. These efforts include Mayo Clinic's more than $9 billion investment to invent a new integrated health care experience. This transformation includes bringing the first carbon ion therapy program in the Americas to the Mayo Clinic and advanced biomanufacturing capabilities to all Mayo Clinic destination medical center sites, among other advances. Dr. Farrugia has been a Mayo Clinic physician for more than 35 years. He is jointly appointed in the Division of Gastroenterology and Hepatology, the Department of Internal Medicine, and the Department of Physiology and Biomedical Engineering. He is also a professor of medicine and physiology at the Mayo Clinic College of Medicine and Science. He completed his undergraduate training at St. Aloysius College and earned his medical degree from the University of Malta Medical School. (**Conflict and Interest Disclosures:** None)

Kadija Ferryman is an anthropologist who studies equity, ethics, and policy in health risk technologies. Dr. Ferryman is faculty at the Berman Institute of Bioethics and Assistant Professor in the Department of Health Policy and Management at the Johns Hopkins Bloomberg School of Public Health. Before her training as an anthropologist, Dr. Ferryman began her professional career as a policy researcher at the Urban Institute in Washington, DC. She has published in journals such as *JAMA*, the *New England Journal of Medicine*, the *American Journal of Bioethics*, and the *Journal of the American Informatics Association*. Dr. Ferryman received her B.A. in anthropology from Yale University and her Ph.D. in anthropology from the New School for Social Research. (**Conflict and Interest Disclosures:** Member, Institutional Review Board, All of Us Research Program, National Institutes of Health; Member, Digital Ethics Advisory Panel, Merck KGaA [Merck Germany])

Sanjay Gupta is the multiple Emmy® Award–winning chief medical correspondent for CNN and host of the CNN podcast "Chasing Life." Dr. Gupta, a practicing neurosurgeon, plays an integral role in CNN's reporting on health and medical news for all CNN shows domestically and internationally, and regularly contributes to CNN Digital. Since 2001, Dr. Gupta has covered some of the most important health stories in the United States and around the world. On March 9, 2020, Dr. Gupta penned an op-ed announcing the network would refer to the novel coronavirus outbreak as a "pandemic," ahead of the World Health Organization and the Centers for Disease Control and Prevention. Throughout 2020 into 2021, Dr. Gupta reaffirmed his role as a trusted guide to viewers worldwide on navigating between facts and fiction surrounding COVID-19 and the pandemic. In addition to his work for CNN, Dr. Gupta is an associate professor of neurosurgery at Emory University Hospital and the associate chief of neurosurgery at Grady Memorial Hospital in Atlanta. He serves as a diplomat of the American Board of Neurosurgery. In 2019, Dr. Gupta was elected to the National Academy of Medicine. His upcoming book, *It Doesn't Have to Hurt: Your Smart Guide to a Pain-Free Life*, will be published in September 2025 with Simon & Schuster. Dr. Gupta received his undergraduate degree from the University of Michigan and M.D. from the University of Michigan Medical School. (**Conflict and Interest Disclosures:** None)

Eric Horvitz serves as Microsoft's chief scientific officer, guiding strategic scientific initiatives companywide. He serves on the National Institutes of Health's (NIH's) Advisory Committee to the Director working group on artificial intelligence (AI) and previously served on the President's Council of Advisors on Science and Technology, co-leading studies in health care and biosciences. Dr. Horvitz has also served on the Board of Regents of the National Library of Medicine, as commissioner on the National Security Commission on AI, and as the president of the Association for the Advancement of AI (AAAI). Dr. Horvitz's research and contributions span machine learning (ML), reasoning, and human–AI interaction. For decades, he has advanced AI applications in health care settings, including harnessing ML to develop and field diagnostic and predictive models. His foundational contributions in AI include probabilistic and decision-theoretic reasoning, models of bounded rationality, and human–AI complementarity and coordination. His honors include the Feigenbaum Prize and the Allen Newell Prize for his contributions in AI; induction into the CHI Academy for advances in human–AI collaboration; election as a member of the National Academy of Engineering, the American Academy of Arts & Sciences, and the American Philosophical Society; and as a fellow of AAAI, the Association for Computing Machinery, the Association for the Advancement of Science, and the American College of Medical Informatics. He co-founded the Partnership on AI and Stanford's One Hundred Year Study on AI. At Microsoft, he established the Aether Committee, advising on AI safety, trustworthiness, and ethics, and co-founded the Office of Responsible AI, overseeing policies and compliance across company products and services. (**Conflict and Interest Disclosures:** Employment by Microsoft Corporation; Member, NIH Advisory Committee to the Director Working Group on AI)

Roy Jakobs is the chief executive officer (CEO) of Royal Philips. As the CEO, he is also chair of the Board of Management and the Executive Committee. With his extensive global executive leadership experience, Jakobs drives Philips's strategy to help deliver better care for more people. He is committed to helping health care professionals provide better care in hospitals, clinics, and the home and empowering people to take care of their health and well-being. His track record over the past 25 years reflects his passion to help address wider societal challenges. He has driven (digital) transformations in energy, scientific information publishing, and health technology in multinational companies. In doing so, he has built high-performing teams, leveraged digital innovation and mergers and acquisitions in business-to-business, business-to-consumer, and business-to-government segments to create value for multiple stakeholders, including patients, customers, and society. Jakobs has an M.B.A. from Radboud University Nijmegen and the Università degli Studi di Bologna, Italy. He also has a master's degree in marketing from the TIAS School for Business and Society and completed the New Board Program from Nyenrode Business University, both in the Netherlands. (**Conflict and Interest Disclosures**: None)

Kevin B. Johnson is the David L. Cohen University Professor of Biomedical Informatics, Computer Science, Pediatrics, and Science Communication at the University of Pennsylvania, and the vice president of applied clinical informatics in the University of Pennsylvania Health System. He received his M.D. from Johns Hopkins University and his M.S. in medical informatics from Stanford University. Dr. Johnson is an internationally respected expert in clinical informatics. His current research focuses on clinical documentation and artificial intelligence. Dr. Johnson was among the world's first researchers to propose and demonstrate the value of text messaging in behavior change. Previously, he served as the chair of the Department of Biomedical Informatics and the chief informatics officer for the Vanderbilt University Medical Center. Dr. Johnson holds numerous national leadership positions and serves on various advisory boards. Dr. Johnson is passionate about educating lay audiences about informatics. He has produced documentaries related to health information technology. His podcast "Informatics in the Round" features experts discussing informatics topics to songwriters. Most recently, he co-published a book series called *Who, Me?* featuring scientists from marginalized groups, encouraging young children to consider careers in science, technology, engineering, mathematics, and medicine. He has authored more than 200 publications and has won numerous national awards. He was elected to the American College of Medical Informatics in 2004 and the Academic Pediatric Society in 2010, the National Academy of Medicine in 2010, the International

Association of Health Science Informatics in 2021, and the American Institute of Medical and Biological Engineering in 2022. (**Conflict and Interest Disclosures**: Member, National Scientific Advisory Board, University of Nebraska, Child Health Research Institute; National Advisory Committee Member, Robert Wood Johnson Foundation; Scientific Advisory Board, University of Michigan Taubman Institute; External Advisory Board, Washington University School of Medicine, Institute of Informatics; JAMA Health Forum Editorial Board; Past President, American College of Medical Informatics)

Peter Lee is the president of Microsoft Research. He leads Microsoft Research across its 11 worldwide laboratories to advance human knowledge and incubate research-powered products in artificial intelligence (AI), computer science, health, and life sciences. Before joining Microsoft in 2010, he established a new technology office at the Defense Advanced Research Projects Agency, creating operational capabilities in machine learning, data science, and computational social science. Prior to that, he was a professor and the head of the Computer Science Department at Carnegie Mellon University. Dr. Lee is a member of the National Academy of Medicine and serves on the boards of several institutions in AI and medicine, including the Board of Trustees of the Mayo Clinic and the Boards of Directors of the Kaiser Permanente School of Medicine and the Brotman Baty Institute for Precision Medicine. He served on President Obama's Commission on Enhancing National Cybersecurity and has testified before both the U.S. House Science and Technology and the U.S. Senate Commerce Committees. With Carey Goldberg and Dr. Isaac Kohane, he is the co-author of the book *The AI Revolution in Medicine: GPT-4 and Beyond*. In 2024, Dr. Lee was named by *Time* magazine as one of the 100 most influential people in health and life sciences. (**Conflict and Interest Disclosures:** Employment by Microsoft Corporation)

Kenneth Mandl directs the Computational Health Informatics Program at Boston Children's Hospital and is the Donald A.B. Lindberg Professor of Pediatrics and Biomedical Informatics at Harvard Medical School. He is trained as a pediatrician and pediatric emergency physician. Dr. Mandl is a member of the National Academy of Medicine and as part of its Leadership Consortium, co-chairs its Digital Health and AI Action Collaborative. He has had a sustained influence on the field of biomedical informatics, innovating in biosurveillance, federated data sharing, patient control of data, and health care interoperability. Dr. Mandl's advancements in SMART programming interfaces, in conjunction with his influence on the 21st Century Cures Act, have streamlined universal access to individual and population health data. These capabilities enhance interoperability in health care systems and foster substantial economies of scale. He leads, across seven children's hospitals, the Genomic Information Commons and directs the PrecisionLink Biobank for Health Discovery at Boston Children's Hospital. Dr. Mandl was elected to the American Society for Clinical Investigation, the Society for Pediatric Research, the American College of Medical Informatics, and the American Pediatric Society. He is a recipient of the Presidential Early Career Award for Scientists and Engineers, the Donald A.B. Lindberg Award for Innovation in Informatics, and the Clifford A. Barger Award for top mentors at Harvard Medical School. He was the advisor to two directors of the Centers for Disease Control and Prevention and chaired the Board of Scientific Counselors of the National Institute of Health's National Library of Medicine. (**Conflict and Interest Disclosures**: Boston Children's Hospital receives philanthropic contributions on behalf of Dr. Mandl's laboratory from the SMART Advisory Committee with members including Cambia, Humana, HCA Healthcare, and BMJ Group. Equity in SMART Check-In.)

Kedar Mate is the chief medical officer and the co-founder of Qualified Health. He is the former president and chief executive officer of the Institute for Healthcare Improvement (IHI), a global organization advancing equitable health outcomes through improvement science. With more than 20 years of experience in health care management, public health, and quality improvement, he champions innovative approaches to enhance global health. Dr. Mate co-hosts the "Turn On The Lights" podcast, where he delves into the intersection of health, social justice, and leadership with a diverse array of inspiring guests. He also co-leads the Rise to Health Coalition, a national health equity initiative by IHI, the American Medical Association, and Race Forward. Dr. Mate is passionate about creating systems that address social determinants of health, foster collaboration and learning, and promote innovation and excellence. (**Conflict and Interest Disclosures:** Co-founder and chief medical officer of a generative AI health care company, Qualified Health)

Deven McGraw is the chief regulatory and privacy officer for Citizen Health, a platform for patients to gather, manage, and share their complete health histories (previously known as Ciitizen and recently divested from Invitae). From 2015–2017, she directed U.S. health privacy and security as the deputy director of health information privacy at the Department of Health and Human Services' Office for Civil Rights and as the chief privacy officer (acting) of the Office of the National Coordinator for Health Information Technology. Widely recognized for her expertise in health privacy, she directed the Health Privacy Project at the Center for Democracy & Technology for 6 years, testifying before Congress on health privacy issues on multiple occasions. She currently serves on the federal Health Information Technology Advisory Committee, the Steering Committee for Carequality, the California Data Sharing Agreement Policies and Procedures Subcommittee, and the Data and Surveillance Workgroup of the Center for Disease Control and Prevention's (CDC's) Advisory Committee to the director on CDC's Data Modernization Initiative. She also is on the board of Manifest MedEx, the largest health information exchange in California. She previously was the chief operating officer of the National Partnership for Women and Families and, before joining federal government service, advised health industry clients on Health Insurance Portability and Accountability Act compliance and data governance while a partner at Manatt, Phelps & Phillips, LLP. Deven graduated magna cum laude from Georgetown University Law Center and has an M.P.H. from the Johns Hopkins Bloomberg School of Public Health. (**Conflict and Interest Disclosures**: Employment and minority shareholder interest in Citizen Health, Inc.)

Bakul Patel is the senior director of global digital health strategy and regulatory at Google, focused on building a unified digital health strategy that is aligned with evolving global regulatory needs. Patel's vision is to help realize the potential of technology and its role in democratizing access to high-quality, equitable health care. Prior to joining Google, Patel held the position of the chief digital health officer of global strategy and innovation and the founding director for the Digital Health Center of Excellence at the U.S. Food and Drug Administration (FDA). In these roles, he provided thought leadership and expertise and shaped responsible regulation that balanced innovation and patient safety for digital health. Patel coined the term "software as a medical device" and authored a risk framework and playbook that is now adopted by many medical device regulators globally. He was also the architect of the software precertification pilot program and FDA's framework for artificial intelligence/machine learning–based software that created the predetermined change control plan novel approach for FDA. Patel earned an M.S. in electronic systems engineering from the University of Regina, Canada, and an M.B.A. in international business from Johns Hopkins University. (**Conflict and Interest Disclosures**: Employment by Google, LLC)

Philip Payne serves as the chief health artificial intelligence (AI) officer for BJC Healthcare and Washington University Medicine and he is the founding director of their joint Center for Health AI. He also holds the Janet and Bernard Becker Professorship and is the founding director of the Institute for Informatics, Data Science, and Biostatistics at the Washington University School of Medicine. Additionally, Dr. Payne is a professor of general internal medicine and computer science and engineering. With more than 300 publications, he leads a dynamic research group that addresses areas such as (1) AI-driven methods for discovering and analyzing biomolecular and clinical phenotypes, (2) interventional applications of electronic health records and clinical decision support, (3) human factors and workflow optimization in health care information technology, and (4) the development and assessment of data sharing and analytics platforms to support high-value, agile health care systems and research initiatives. Dr. Payne is an elected fellow of the American College of Medical Informatics, the American Medical Informatics Association (AMIA), the American Institute for Medical and Biological Engineering, and the International Academy of Health Sciences Informatics. He currently serves as the president-elect of AMIA, the leading professional organization in biomedical and health informatics. Beyond academia, Dr. Payne is an active entrepreneur, founding multiple digital health companies and serving in advisory and governance roles with various health and life sciences companies and venture capital firms. (**Conflict and Interest Disclosures**: None)

Vardit Ravitsky is the president and chief executive officer of the Hastings Center, an independent, non-partisan bioethics research institute that is among the most prestigious bioethics and health policy institutes in the world. She is a senior lecturer on global health and social medicine at Harvard Medical School, and was a full professor at the Bioethics Program, School of Public Health, University of Montreal. She is the past-president of the International Association of Bioethics and a fellow of the Canadian Academy of Health Sciences and of The Hastings Center. Dr. Ravitsky's research focuses on the ethics of genomics and reproduction and the use of artificial intelligence in biomedical research. It is funded by Canada's leading funding agencies and the National Institutes of Health (NIH). She has published more than 200 articles and commentaries on bioethical issues and given more than 300 talks worldwide and more than 400 media interviews. Dr. Ravitsky holds a B.A. from the Sorbonne University in Paris, an M.A. from the University of New Mexico in the United States, and a Ph.D. from Bar-Ilan University in Israel. Previously, she was a fellow in the Department of Bioethics at NIH and faculty at the Department of Medical Ethics, School of Medicine, University of Pennsylvania. She was also a senior policy advisor at the Canadian Institutes of Health Research's Ethics Office and a GE3LS (genomics and its ethical, economic, environmental, legal, and social implications) consultant to Genome Canada. (**Conflict and Interest Disclosures**: None)

Suchi Saria holds a John C. Malone endowed chair and is the director of the Machine Learning and Healthcare Lab at Johns Hopkins University, where she is jointly appointed as faculty in computer science, medicine, and health policy. She is also the founder of Bayesian Health, a clinical artificial intelligence (AI) platform company spun out of Johns Hopkins that augments care teams by bringing together the state of the AI and machine learning (ML) technology combined with responsible AI best practices to dramatically improve quality while saving clinicians' time. Dr. Saria's work in AI over the past two decades has led to foundational advances in the technology, best practices around translation, and AI policy. She has written several seminal papers in AI/ML around issues of learning robust models, detecting drifts, and monitoring and learning from messy real-world datasets. Her applied research has built on these technical advances to develop novel next-generation diagnostic and treatment planning tools that use AI/ML to individualize care. Her work has been funded by leading organizations including the National Science Foundation, the Defense Advanced Research Projects Agency, the U.S. Food and Drug Administration, the National Institutes of Health, and the Centers for Disease Control and Prevention and she regularly serves as a scientific advisor to leading Fortune 500 companies. Dr. Saria completed her Ph.D. in AI at Stanford University. In 2024, she received an honorary doctorate from Mount Holyoke. She is a Sloan Research Fellow, named by the Institute of Electrical and Electronics Engineers to "AI's 10 to Watch," Modern Healthcare's Top 25 Innovators, World Technology Forum's Technology Pioneer, and her work was recognized as one of *Time* Magazine's Best Inventions in 2023 and 2024. (**Conflict and Interest Disclosures**: Stock for scientific and technical work in Bayesian Health, Duality Tech, Century Health, Midstream Health, and Latent; Grant/Research support from the National Science Foundation, National Institutes of Health, and Food and Drug Administration)

Eric Topol is the executive vice-president of Scripps Research, the largest non-profit biomedical research in the United States, where he was the founder and directs the Scripps Research Translational Institute as the chair and a professor of translational medicine. He has published more than 1,300 peer-reviewed articles, with more than 370,000 citations. He was elected to the National Academy of Medicine and is one of the top 10 most cited researchers in medicine. His research is dedicated to promoting human health using genomic, digital and artificial intelligence. He authored three best-seller books on the future of medicine: *The Creative Destruction of Medicine*, *The Patient Will See You Now*, and *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. His new book is *Super Agers: An Evidence-Based Approach to Longevity*. Dr. Topol is the principal investigator to two large National Institutes of Health grants, the All of Us Research Program that supports precision medicine and a Clinical and Translational Science Award that promotes innovation in medicine. He was the founder of a new medical school at Cleveland Clinic (Lerner College of Medicine), was commissioned by the United Kingdom to lead a review of its National Health Service, and is active clinically as a cardiologist. At Substack, he publishes "Ground Truths," a weekly newsletter and podcast on cutting-edge biomedical advances. (**Conflict and Interest Disclosures**: Advisor to Abridge, Tempus Labs, and Pheno.ai)

Selwyn Vickers is the president and chief executive officer (CEO) of Memorial Sloan Kettering Cancer Center (MSK) and the director of the MSK Comprehensive Cancer Center. He is a world-renowned surgeon, pancreatic cancer researcher, and pioneer in health disparities research. His major research interests include gene therapy as an application in the treatment of pancreatobiliary tumors, the role of growth factors and receptors in the oncogenesis of pancreatic cancer, the implications of FAS expressions and Tamoxifen in the growth and treatment of cholangiocarcinoma, assessment of clinical outcomes in the surgical treatment of pancreatobiliary tumors, and the role of death receptors in the treatment of pancreatic cancer. Dr. Vickers is a member of the National Academy of Medicine (NAM), a fellow of the American Association for Cancer Research, and an honorary member of the American Society for Clinical Investigation. He is a member of more than 22 professional societies and has held leadership roles in many, including the NAM, the American Surgical Association, and the Society of Black Academic Surgeons. He has more than 270 peer-reviewed publications spanning 25 years of National Institutes of Health funding investigating the molecular basis for pancreatic adenocarcinoma, developing novel therapeutic approaches, and understanding disparities in cancer incidence, access to care, and clinical trial enrollment. Dr. Vickers has held several academic leadership roles, including the chair of the Department of Surgery at the University of Minnesota Medical School, the dean and the senior vice president of the Heersink School of Medicine at the University of Alabama at Birmingham (UAB), and CEO of the UAB Health Systems and CEO of the UAB/Ascension St. Vincent's Alliance. (**Conflict and Interest Disclosures**: None)

EDITOR INFORMATION

Laura Adams is a senior advisor at the National Academy of Medicine (NAM). She provides strategic counsel and leadership for the Science and Technology portfolio of the NAM Leadership Consortium and its initiatives on the digital infrastructure and accelerated use of artificial intelligence (AI) in health, health care, and biomedical science. She leads the NAM's AI Code of Conduct (AICC) national initiative and her expertise is in AI, digital health, and human-centered care. Adams is a member of the international AI Expert Panel for the Organisation of Economic Co-operation and Development, the Bipartisan Policy Center's Health Care AI Advisory Panel, and the Consumer Technology Association's Health AI Planning Council. Adams chairs the Global Opportunities Group for The Centre of Excellence for AI Regulatory Science and Innovation in the United Kingdom. She is a strategic advisor for Maverick Health Policy in Washington, DC, and Inflammatix, a Burlingame, California-based biotech company specializing in host immune response diagnostics. Prior to her work at the NAM, Adams was the founding president and chief executive officer of the Rhode Island Quality Institute, Rhode Island's statewide health information exchange. (**Conflict and Interest Disclosures**: Board member, TMA Precision Health; Strategic Advisor, Inflammatix; Strategic Advisor, Maverick Health Policy)

Elaine Fontaine currently serves as a special advisor to the Science and Technology portfolio of the Leadership Consortium at the National Academy of Medicine. With more than 30 years of senior leadership experience in health information technology across provider, payer, research, and health information exchange sectors, she is recognized for translating emerging technologies into practical, real-world solutions. Prior to her consulting role, she was the chief operating officer at the Rhode Island Quality Institute (RIQI), overseeing operations in finance, information technology, analytics, compliance, contracting, human resources, sales, and grant management. During her tenure, RIQI received several notable accolades, including the 2017 Innovation Award in Healthcare in Rhode Island, the 2018 national Healthcare Informatics Innovation Award for its impact on the opioid crisis, and recognition as a finalist for the 2018 New England Business Innovation Award. Fontaine has also co-authored multiple peer-reviewed publications and presented her work at both local and national technology and analytics forums. (**Conflict and Interest Disclosures**: None)

Sunita Krishnan serves as a senior program officer on the National Academy of Medicine's (NAM) Leadership Consortium team. Krishnan leads the Science and Technology portfolio, which houses the Digital Health and AI Action Collaborative and the Evidence and Data Action Collaborative. In addition, she serves as the project manager for the AI Code of Conduct project. Prior to joining the NAM, Krishnan served as a senior manager at AcademyHealth. She managed the day-to-day activities of the Evidence-Informed State Health Policy Institute that aims to increase the use of relevant, timely, and translatable evidence in state policymaking to improve health and health care quality, outcomes, equity, accessibility, and affordability. Prior to joining AcademyHealth, Krishnan worked as a health policy research assistant for the Healthcare Value Hub, where she wrote research reports on various health value and quality topics, such as telehealth, patient-shared decision making, pharmaceutical costs, and more. She also worked at the American Association for Retired Persons (AARP) as a legislative fellow where she provided research support on state legislation pertaining to long-term care and logistical support for meetings. Krishnan graduated with a B.S. in biology and public health from the University of Minnesota and received her M.P.H. specializing in health policy from The George Washington University. (**Conflict and Interest Disclosures**: None)

Michael Matheny is the director for the Center for Improving the Public's Health with Informatics and a professor in the Departments of Biomedical Informatics, Medicine, and Biostatistics at the Vanderbilt University Medical Center. He is also a part-time primary care physician, physician scientist, board certified in internal medicine and clinical informatics, and the associate director of health systems research and development in the VA Informatics and Computing Infrastructure at the Tennessee Valley Healthcare System VA in Nashville, Tennessee. He is an elected fellow of the American College of Medical Informatics, an elected fellow of the American Medical Informatics Association, and an elected member of the American Society of Clinical Investigation. Dr. Matheny's work has focused on developing and adapting artificial intelligence and machine learning methods for medical product active surveillance, algorithm vigilance, probabilistic phenotyping, natural language processing, and risk prediction modeling. (**Conflict and Interest Disclosures**: None)

CONTRIBUTOR INFORMATION

David Dorr is an internal medicine doctor and the chief research information officer at Oregon Health & Science University. He focuses on improving capabilities and the use of innovations to manage data, information, and knowledge in research and in translating it to health care. He earned his bachelor's degree in economics (with minors in mathematics and psychology) and his M.D. from Washington University in St. Louis. He then completed his internal medicine residency at OHSU and earned a master's degree in medical informatics and health services administration from the University of Utah. Broadly, Dr. Dorr's interests lie in complex care management, especially for older adults and other at-risk populations, coordination of care, collaborative care, chronic disease management, quality, and the requirements of clinical information systems to support these areas. From these interests, he has broadened into clinical information needs, electronic health record deployment and health information exchange as a way to expand systems-based approaches to all of health care. Finally, Dr. Dorr performs evaluations of care management and informatics initiatives using a variety of methodologies. (**Conflict and Interest Disclosures:** None)

Andrea Downing is a security researcher and advocate. She has expressed significant concerns about health privacy, particularly in the context of online patient communities and the use of social media for support groups. Her work emphasizes the vulnerabilities and risks associated with the sharing and handling of sensitive health information in digital spaces. Downing's advocacy is rooted in her personal journey as a BRCA1 mutation "previvor," which has driven her to focus on privacy protection within the breast cancer community. She co-founded The Light Collective, aiming to create safe spaces on the internet where individuals can share their experiences and receive support without compromising their privacy. (**Conflict and Interest Disclosures:** Board President of The Light Collective)

Tyler Loftus is a trauma and acute care surgeon and an intensive care unit doctor in the University of Florida Department of Surgery. He treats patients for traumatic injury and those needing emergency general surgery. He is passionate about resident education and professional development and serves as the program director for the University of Florida General Surgery residency. His National Institutes of Health–funded research pursuits have evolved from translational science, focusing on bone marrow failure and anemia after traumatic injury, to data science, focusing on machine learning to augment personalized, patient-centered decision making. (**Conflict and Interest Disclosures:** None)

Shauna Overgaard works to advance the safe, effective, and equitable integration of artificial intelligence (AI) into clinical practice through the development of enterprise frameworks for evaluation, implementation, and oversight. Her efforts center on aligning innovation with regulatory science, implementation science, and clinical usability to ensure AI technologies meet the highest standards of patient care and accelerate innovation to practice. Dr. Overgaard is the senior director of AI strategy and frameworks at the Mayo Clinic and the co-director of the AI Validation & Stewardship Program. Her scholarly work focuses on translational AI governance, clinical assurance, and equitable adoption. She serves on the Editorial Board of npj Health Systems and as a special topics guest editor for *BMJ Health & Care Informatics* and the *Mayo Clinic Proceedings: Digital Health*. Dr. Overgaard is recognized as a Rising Star in Modern Healthcare's 2025 Leading Women in Healthcare. Before her work in AI translation, she was involved in clinical research focused on the development of diagnostic tests leveraging principles of graph theory and multimodal data in the realm of neuroimaging, proteomic, and genomic data. She continues to maintain a responsibility to help facilitate national consensus-building efforts related to medical informatics and health AI. (**Conflict and Interest Disclosures:** None)

Ravi B. Parikh is an oncologist and an associate professor at Emory University and the Winship Cancer Institute. His research investigates how doctors, patients, and health policymakers can use artificial intelligence (AI) to improve decision making and promote equitable health care. Dr. Parikh's lab, the Human-Algorithm Collaboration Lab, runs clinical trials testing how clinicians and AI collaborate in making decisions about diagnosis, prognosis, and treatment selection. Dr. Parikh's work has been published in *Science*, the *New England Journal of Medicine*, and *JAMA* and he has written about AI for *The New York Times* and *The Washington Post*. He sits on the Board of the Coalition to Transform Advanced Care. Dr. Parikh graduated from Harvard Medical School and the Kennedy School of Government. He completed his clinical training at Brigham and Women's Hospital and the University of Pennsylvania. Dr. Parikh has received awards from the National Academy of Medicine, the American Medical Association, the American Society for Clinical Oncology, and the American College of Physicians. (**Conflict and Interest Disclosures:** Grants from various foundations and other non-commercial organizations for research on AI in health and health care decision making; personal fees and equity from GNS Healthcare, Thyme Care, Main Street Health, and Onc. AI; personal fees from the ConcertAI, Cancer Study Group, Mendel.ai, Optinosis, Biofourmis, Archetype Therapeutics, CreditSuisse, G1 Therapeutics, Humana, and Nanology; honoraria from Flatiron and Medscape; board membership [unpaid] at the Coalition to Transform Advanced Care and American Cancer Society; editor at the *Journal of Clinical Oncology*; and serves on a leadership consortium [unpaid] at the National Quality Forum, all outside the submitted work)